

(19)



Евразийское  
патентное  
ведомство

(21)

201791414

(13)

A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки  
2018.02.28

(51) Int. Cl. C12N 15/82 (2006.01)  
C12N 15/29 (2006.01)  
C07K 14/415 (2006.01)  
A01H 1/08 (2006.01)  
A01H 5/00 (2006.01)  
A01H 5/06 (2006.01)  
A01H 5/10 (2006.01)

(22) Дата подачи заявки  
2015.12.23

(54) ГАПЛОИДНЫЙ ИНДУКТОР

(31) 14004389.4

(32) 2014.12.23

(33) ЕР

(86) PCT/EP2015/081158

(87) WO 2016/102665 2016.06.30

(88) 2017.03.16

(71) Заявитель:

КВС ЗААТ СЕ (DE)

(72) Изобретатель:

Болдуан Кристоф, Бройер Франк,  
Клойбер-Майтц Моника, Ниссен  
Маркус, Оузунова Милена, Шульц  
Бритта, Викторст Силке (DE)

(74) Представитель:

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

(57) Настоящее изобретение относится к нетрансгенным и трансгенным растениям, предпочтительно зерновым культурам, обладающим биологической активностью гаплоидного индуктора и содержащим полинуклеотид, включающий нуклеотидную последовательность, кодирующую центромерный гистоновый H<sub>3</sub> (CENH<sub>3</sub>) белок, при этом полинуклеотид содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности белка CENH<sub>3</sub>, а также какой-либо аминокислоты, входящей в его состав. Кроме того, настоящее изобретение обеспечивает способы создания растений-индукторов, способы создания гаплоидных растений и удвоенных гаплоидов на основании использования растений-индукторов, а также способы облегчения цитоплазматического обмена.

201791414

A1

A1

201791414

## ГАПЛОИДНЫЙ ИНДУКТОР

Настоящее изобретение относится к нетрансгенным и трансгенным растениям, предпочтительно к зерновым культурам, которые обладают биологической активностью гаплоидных индукторов и которые содержат по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности белка CENH3, при этом изменение придает биологическую активность гаплоидного индуктора. Кроме того, изобретение касается методов создания растений по данному изобретению, гаплоидных растений и удвоенных гаплоидов, получаемых путем скрещивания растений по данному изобретению с растениями дикого типа, а также методов облегчения цитоплазматического обмена.

Создание и использование гаплоидов является одним из важнейших биотехнологических средств улучшения культивируемых растений. Преимущество гаплоидов для селекционеров заключается в том, что гомозиготность может достигаться уже в первом поколении после дигаплоидизации с образованием удвоенных гаплоидов без необходимости возвратного скрещивания на протяжении нескольких поколений с целью достижения высокой степени гетерозиготности. Кроме того, ценность гаплоидов в исследованиях и селекции растений состоит в том, что клетки-основатели удвоенных гаплоидов являются продуктами мейоза, поэтому получаемые в результате поколения представляют собой пулы различных рекомбинантных и одновременно генетически устойчивых индивидуумов. Таким образом, создание удвоенных гаплоидов обеспечивает не только весьма полезную генетическую вариабельность при отборе с целью улучшения культур, но также является ценным средством для получения картирующихся популяций, рекомбинантных инбредных линий, а также быстрого получения гомозиготных мутантов и трансгенных линий.

Гаплоиды можно получать, используя подходы *in vitro* или *in vivo*. Однако многие виды и генотипы трудно поддаются применению таких подходов. Как альтернатива, существенные изменения центромер-специфического варианта гистона H3 (CENH3, также называемого CENP-A) путем замены его N-терминальных участков и его слияния с зеленым флуоресцентным белком GFP (“GFP-tailswap” CENH3) приводит к образованию линий-индукторов гаплоидов у модельного растения *Arabidopsis thaliana* (Ravi и Chan, Nature, 464 (2010), 615-618; Comai, L,

“Элиминация генома: преобразование фундаментальных исследований в инструмент для будущей селекции растений”, PLoS biology, 12.6 (2014)). Белки CENH3 представляют собой варианты H3 гистоновых белков, входящих в состав кинетохорного комплекса активных центромер. С участием “GFP-tailswap” линий-индукторов гаплоидов в потомстве происходит гаплоидизация при скрещивании растения гаплоиндутора с растением дикого типа. Интересно, что линия-индуктор гаплоидов остается стабильной при самовоспроизводстве, из чего можно предположить, что конкуренция между модифицированной центромерой и центромерой дикого типа в развитии гибридного зародыша приводит к инактивации центромеры родителя-индуктора и, как следствие, элиминации хромосом одного из родителей. В результате хромосомы, содержащие измененный белок CENH3, утрачиваются на ранней стадии развития зародыша и образуется гаплоидное потомство, содержащее хромосомы только родителя дикого типа.

Таким образом, гаплоидные растения можно получать путем скрещивания “GFP-tailswap” трансгенных растений как индуктора гаплоидов с дикорастущими растениями. Однако, как показано выше, эта технология требует значительных изменений белка CENH3, при этом в растения включается гетерологичный трансген, что экономически проблематично ввиду растущего общественного нерасположения к сельскохозяйственным культурам, создаваемым на основе генной инженерии.

В связи с этим одной из целей настоящего изобретения является преодоление вышеупомянутых проблем и, в частности, обеспечение альтернативных растений в качестве индуктора гаплоидов, которые не содержат значительных модификаций белка CENH3 и/или не созданы на основе генной инженерии.

Эта проблема решается при помощи объекта, раскрытого в независимых пунктах формулы изобретения, в частности, при помощи растения, обладающего биологической активностью гаплоиндутора и содержащего полинуклеотид, который включает нуклеотидную последовательность, кодирующую центромерный гистоновый H3 (CENH3) белок, включающий домен CATD, при этом полинуклеотид содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности белка CENH3, которое придает биологическую активность гаплоидного индуктора. В контексте настоящего изобретения термин “изменение” означает любую модификацию аминокислотной последовательности белка CENH3 (включая множественные модификации),

вызываемую по меньшей мере одной мутацией полинуклеотида, включающего нуклеотидную последовательность, кодирующую центромерный гистоновый H3 (CENH3) белок. Полинуклеотид может быть геномной ДНК гена CENH3 или кДНК CENH3, либо 5'- или 3'-нетранслируемыми областями гена CENH3, или их смесью, содержащей, например, часть геномной ДНК и часть кДНК. Изменение может представлять собой замену одной или нескольких аминокислот, инсерцию одной или нескольких аминокислот, либо делецию одной или нескольких аминокислот. Мутации на уровне ДНК, способные привести к изменению аминокислотной последовательности белка CENH3, могут быть точечными мутациями, приводящими к замене аминокислоты или стоп-кодона, инсерциями или делециями, которые сдвигают рамку считывания гена CENH3, либо мутациями в сайтах сплайсинга.

В одном предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену аминокислотной последовательности белка CENH3, которая придает биологическую активность гаплоидного индуктора, в по меньшей мере одном сегменте аминокислотной последовательности белка CENH3. Сегмент последовательности выбирают из группы, включающей N-терминальный хвостовой домен, CATD домен,  $\alpha$ N-спираль,  $\alpha$ 1-спираль, петлю1,  $\alpha$ 2-спираль, петлю2,  $\alpha$ 3-спираль, а также С-терминальный домен. N-терминальный хвостовой домен соответствует аминокислотной последовательности в положении 1-82, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или N-терминальный хвостовой домен белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 1-246, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*. Домен CATD белка CENH3 соответствует аминокислотной последовательности в положении 113-155, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или домен CATD белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 337-465, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*.  $\alpha$ N-спираль белка CENH3 соответствует аминокислотной последовательности в положении 83-97, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или  $\alpha$ N-спираль белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 247-291, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*.  $\alpha$ 1-спираль белка CENH3 соответствует аминокислотной последовательности в положении 103-113, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или  $\alpha$ 1-спираль белка CENH3 кодируется нуклеотидной последовательностью с

нуклеотидами в положении 307-339, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*. Петля1 белка CENH3 соответствует аминокислотной последовательности в положении 114-126, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или петля1 белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 340-378, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*.  $\alpha$ 2-спираль белка CENH3 соответствует аминокислотной последовательности в положении 127-155, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или  $\alpha$ 2-спираль белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 379-465, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*. Петля2 белка CENH3 соответствует аминокислотной последовательности в положении 156-162, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или петля2 белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 466-486, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*.  $\alpha$ 3-спираль белка CENH3 соответствует аминокислотной последовательности в положении 163-172, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или  $\alpha$ 3-спираль белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 487-516, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*. С-терминальный домен белка CENH3 соответствует аминокислотной последовательности в положении 173-178, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, и/или С-терминальный домен белка CENH3 кодируется нуклеотидной последовательностью с нуклеотидами в положении 517-534, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*. Последовательности *A. thaliana* используются только в качестве эталонных и не ограничивают изобретение определенными последовательностями *A. thaliana*. Вследствие высокой степени консервативности последовательностей, специалист в данной области техники может выявить нуклеотидную последовательность и аминокислотную последовательность, соответствующую последовательностям *A. thaliana*, в любом другом растительном материале или виде растения.

Белки CENH3 представляют собой варианты гистоновых белков Н3, входящих в состав кинетохорного комплекса активных центромер, т.е. белковой структуры на хромосомах, к которой прикрепляются волокна веретена во время клеточного деления. В основном белки CENH3 характеризуются вариабельным N-терминальным хвостовым доменом и не образуют жесткую вторичную структуру, а также для них характерен консервативный домен гистоновой складки, состоящий из трех  $\alpha$ -спиральных

участков, называемых  $\alpha$ 1- $\alpha$ 3, которые соединяются двумя петельными сегментами. N-терминальный хвостовой домен в первую очередь подвергается посттрансляционной модификации ферментами. Такие модификации включают метилирование, цитруллинирование, фосфорилирование, сумоилирование, убиквитинирование и АДФ-рибозилирование и воздействуют на функцию регуляции гена CENH3. Внутри домена гистоновой складки локализуется высококонсервативный домен CATD (CENP-A targeting domain), образуемый частями  $\alpha$ 1-спиралей, целой  $\alpha$ 2-спиралью и соединительной петлей1. Консервативный домен CATD необходим для загрузки CENH3 при участии шаперонов и поэтому жизненно важен для локализации его кинетохора и функционирования центромеры. N-терминальный хвостовой домен и домен гистоновой складки соединены между собой  $\alpha$ 2-спиралью.

Авторы настоящего изобретения с удивлением обнаружили, что растение, обладающее способностью продуцировать гаплоидное потомство, т.е. гаплоидный индуктор, можно получать не только путем изменения аминокислотной последовательности консервативного белка CENH3, но также путем изменения аминокислотной последовательности любого другого домена и структурных участков гена CENH3 и белка CENH3. Более того, способность продуцировать гаплоидное потомство может еще более усиливаться путем комбинирования двух или более изменений аминокислотной последовательности белка CENH3 в различных доменах, сегментах или структурных участках белка CENH3. В результате эффективность продуцирования гаплоидов значительно возрастает. Предпочтительно, чтобы это достигалось с помощью трансгенных, а также нетрансгенных методов. Предпочтительными являются нетрансгенные методы, так как дерегулирование генетически модифицированных организмов (ГМО) требует больших затрат, а также в связи с растущим общественным неприятием генетически модифицированных организмов (ГМО) или растений, созданных с помощью ГМО, в частности сельскохозяйственных культур для потребления человеком, и необходимостью решения широкого круга вопросов, связанных с допуском на рынок, включая жесткую оценку на безопасность ГМО-продуктов.

Настоящее изобретение обеспечивает растение, включающее и экспрессирующее белок CENH3, характеризующееся тем, что растение содержит полинуклеотид, включающий нуклеотидную последовательность гена, кодирующего белок CENH3, при этом полинуклеотид содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной

последовательности белка CENH3 в по меньшей мере одном сегменте аминокислотной последовательности белка CENH3, где сегмент выбирают из группы, включающей N-терминальный хвостовой домен, наиболее предпочтительно N-терминальный хвостовой домен, имеющий консенсусную последовательность SEQ ID NO. 1 и SEQ ID NO. 2, соответственно,  $\alpha$ N-спираль, наиболее предпочтительно  $\alpha$ N-спираль, имеющую консенсусную последовательность SEQ ID NO. 3,  $\alpha$ 1-спираль, наиболее предпочтительно  $\alpha$ 1-спираль, имеющую консенсусную последовательность SEQ ID NO. 4, петлю1, наиболее предпочтительно петлю1, имеющую консенсусную последовательность SEQ ID NO. 5,  $\alpha$ 2-спираль, наиболее предпочтительно  $\alpha$ 2-спираль, имеющую консенсусную последовательность SEQ ID NO. 6, петлю2, наиболее предпочтительно петлю2, имеющую консенсусную последовательность SEQ ID NO. 7,  $\alpha$ 3-спираль, наиболее предпочтительно  $\alpha$ 3-спираль, имеющую консенсусную последовательность SEQ ID NO. 8 и С-терминальный домен, наиболее предпочтительно С-терминальный домен, имеющий консенсусную последовательность SEQ ID NO. 9. Изменение аминокислотной последовательности белка CENH3 может придавать биологическую активность гаплоидного индуктора растению. В предпочтительном варианте осуществления настоящее изобретение касается растения, содержащего полинуклеотид, включающий нуклеотидную последовательность, которая кодирует центромерный гистоновый H3 (CENH3) белок, при этом полинуклеотид содержит по меньшей мере одну мутацию и по меньшей мере одна мутация вызывает изменение аминокислотной последовательности белка CENH3 в по меньшей мере одном сегменте аминокислотной последовательности белка CENH3. Сегмент может быть а) N-терминальным хвостовым доменом, кодируемым нуклеотидной последовательностью с нуклеотидами в положении 1-246, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующим аминокислотной последовательности в положении 1-82, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 1-177, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующим аминокислотной последовательности в положении 1-59, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 1-252, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующим аминокислотной последовательности в положении 1-84, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 1-186,

приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующим аминокислотной последовательности в положении 1-62, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 1-186, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующим аминокислотной последовательности в положении 1-62, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющим консенсусную последовательность SEQ ID NO. 1 и SEQ ID NO. 2, б) αN-спиралью, кодируемой нуклеотидной последовательностью с нуклеотидами в положении 247-291, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующей аминокислотной последовательности в положении 83-97, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 178-222, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующей аминокислотной последовательности в положении 60-74, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 253-297, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующей аминокислотной последовательности в положении 85-99, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 187-231, приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующей аминокислотной последовательности в положении 63-77, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 187-231, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующей аминокислотной последовательности в положении 63-77, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющей консенсусную последовательность SEQ ID NO. 3, с) α1-спиралью, кодируемой нуклеотидной последовательностью с нуклеотидами в положении 307-339, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующей аминокислотной последовательности в положении 103-113, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 238-270, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующей аминокислотной последовательности в положении 80-90, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 313-345, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующей аминокислотной последовательности в положении 105-115, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемой нуклеотидной последовательностью с

нуклеотидами в положении 247-279, приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующей аминокислотной последовательности в положении 83-93, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 247-279, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующей аминокислотной последовательности в положении 83-93, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющей консенсусную последовательность SEQ ID NO. 4, d) петлей1, кодируемой нуклеотидной последовательностью с нуклеотидами в положении 340-378, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующей аминокислотной последовательности в положении 114-126, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 271-306, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующей аминокислотной последовательности в положении 91-102, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 346-384, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующей аминокислотной последовательности в положении 116-128, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 280-318, приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующей аминокислотной последовательности в положении 94-106, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 280-318, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующей аминокислотной последовательности в положении 94-106, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющей консенсусную последовательность SEQ ID NO. 5, e) α2-спиралью, кодируемой нуклеотидной последовательностью с нуклеотидами в положении 379-465, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующей аминокислотной последовательности в положении 127-155, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 307-393, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующей аминокислотной последовательности в положении 103-131, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 385-471, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующей аминокислотной последовательности в положении 129-157, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемой

нуклеотидной последовательностью с нуклеотидами в положении 319-405, приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующей аминокислотной последовательности в положении 107-135, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 319-405, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующей аминокислотной последовательности в положении 107-135, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющей консенсусную последовательность SEQ ID NO. 6, f) петлей2, кодируемой нуклеотидной последовательностью с нуклеотидами в положении 466-486, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующей аминокислотной последовательности в положении 156-162, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 394-414, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующей аминокислотной последовательности в положении 132-138, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 472-492, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующей аминокислотной последовательности в положении 158-164, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 406-426, приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующей аминокислотной последовательности в положении 136-142, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 406-426, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующей аминокислотной последовательности в положении 136-142, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющей консенсусную последовательность SEQ ID NO. 7, g) α3-спиралью, кодируемой нуклеотидной последовательностью с нуклеотидами в положении 487-516, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующей аминокислотной последовательности в положении 163-172, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 415-444, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующей аминокислотной последовательности в положении 139-148, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 493-522, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующей аминокислотной последовательности в положении 165-

174, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 427-456, приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующей аминокислотной последовательности в положении 143-152, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемой нуклеотидной последовательностью с нуклеотидами в положении 427-456, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующей аминокислотной последовательности в положении 143-152, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющей консенсусную последовательность SEQ ID NO. 8, либо h) С-терминальным доменом, кодируемым нуклеотидной последовательностью с нуклеотидами в положении 517-534, приведенной в SEQ ID NO. 10 из *Arabidopsis thaliana*, и соответствующим аминокислотной последовательности в положении 173-178, приведенной в SEQ ID NO. 11 из *Arabidopsis thaliana*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 445-462, приведенной в SEQ ID NO. 22 из *Beta vulgaris*, и соответствующим аминокислотной последовательности в положении 149-154, приведенной в SEQ ID NO. 23 из *Beta vulgaris*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 523-540, приведенной в SEQ ID NO. 13 из *Brassica napus*, и соответствующим аминокислотной последовательности в положении 175-180, приведенной в SEQ ID NO. 14 из *Brassica napus*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 457-471, приведенной в SEQ ID NO. 19 из *Zea mays*, и соответствующим аминокислотной последовательности в положении 153-157, приведенной в SEQ ID NO. 20 из *Zea mays*, либо кодируемым нуклеотидной последовательностью с нуклеотидами в положении 457-471, приведенной в SEQ ID NO. 16 из *Sorghum bicolor*, и соответствующим аминокислотной последовательности в положении 153-157, приведенной в SEQ ID NO. 17 из *Sorghum bicolor*, либо имеющим консенсусную последовательность SEQ ID NO. 9. Часть  $\alpha$ 1-спирали, полная петля1 и полная  $\alpha$ 2-спираль локализуются внутри домена CATD белка CENH3, как показано выше.

Немутированный N-терминальный хвостовой домен белка CENH3 является частично консервативным у видов растений (см. Рис.). В настоящем изобретении любое положение аминокислоты, касающееся этих двух консервативных частей N-терминального хвостового домена (часть А и часть В) или нижеописываемой консенсусной последовательности, соответствует следующей системе нумерации. Консервативные часть А и часть В могут разделяться одной или несколькими аминокислотами. Конкретный номер

изменяется в зависимости от вида растения. Для этого в консенсусной последовательности введена метка “\*” в качестве указателя месторасположения. Предпочтительно, чтобы немутуированный N-терминальный хвостовой домен демонстрировал аминокислотную последовательность, как показано в Таблице 1.

**Таблица 1.** Специфические аминокислоты, входящие в состав N-терминального хвостового домена белка CENH3

Консервативная часть - положение в N-конце	Аминокислота(ы)
A / 1	M
A / 2	A
A / 3	R
A / 4	T, V, I или A
A / 5	K или R
A / 6	H, T, Q или K
A / 7	X
A / 8	X
A / 9	V, A, P, G, N, P, R, S или H
A / 10	T, R, S, L, K, H, N, A или P
A / 11	R, K, A, N или T
A / 12	S, A, T, L, K, R, D, N или E
A / 13	Q, T, R, A, P, S, G, N, V, K или R
A / 14	P, T, D, E, Q, S, N, G, A, R или R
A / 15	R, N, H, V, G, K, S, A, T, E или P
B / 1	R, D, K, V, G, P, S, Q, T или A
B / 2	G, A, S, K, R, V, T, P или Q
B / 3	S, T, K, V, R, Q, A, E, G, P или D
B / 4	Q, P, N, T, E, K, G, S, R, A или D
B / 5	K, Q, P, G, N, T, H или R
B / 6	X
B / 7	K, R, Q или H
B / 8	K, Q или R
B / 9	S, A, T, K, P или R
B / 10	Y, F, H, T, K, R, F или Q
B / 11	R
B / 12	Y, R, W, F, L, N или S
B / 13	R или K
B / 14	P, A или S

Наиболее предпочтительно, чтобы N-терминальный хвостовой домен содержал консенсусные последовательности SEQ ID NO. 1 (часть А, до \*) и SEQ ID NO. 2 (часть В, за пределами \*), а именно:

MARTK HXXAR RSRKR \* QSQTQ XKKKH RYRP.

5            10            15            5            10            14

Как показано выше, N-терминальный хвостовой домен содержит в своем составе неспецифические [обозначенные символом X] и специфические аминокислоты [обозначенные однобуквенным символом]. Вместо неспецифической аминокислоты "X" может быть выпадение по меньшей мере одной аминокислоты.

Немутуированная  $\alpha$ N-спираль белка CENH3 является высококонсервативной у видов растений и состоит из 15 аминокислот, начиная с аминокислоты в положении 1 и заканчивая аминокислотой в положении 15. В настоящем изобретении любая ссылка на положение аминокислоты, касающееся  $\alpha$ N-спирали или нижеописываемой консенсусной последовательности SEQ ID NO. 3, следует этой системе нумерации. Предпочтительно, чтобы немутуированная  $\alpha$ N-спираль демонстрировала аминокислотную последовательность, как показано в Таблице 2.

**Таблица 2.** Специфические аминокислоты, входящие в состав  $\alpha$ N-спирали белка CENH3

Положение в $\alpha$ N-спирали	Аминокислота(ы)
1	G
2	T
3	V
4	A
5	L
6	K, W или R
7	E или Q
8	I
9	R
10	X
11	F, Y или L
12	Q или R
13	K
14	Q, S или T
15	T, F, W, V, C или A

Наиболее предпочтительно, чтобы  $\alpha$ N-спираль содержала консенсусную последовательность SEQ ID NO. 3, а именно:

GTVAL REIRX FQKTT.

5            10            15

Как показано выше,  $\alpha$ N-спираль содержит в своем составе неспецифические [обозначенные символом X] и специфические аминокислоты [обозначенные однобуквенным символом].

Немутуированная  $\alpha$ 1-спираль белка CENH3 является консервативной у видов растений и состоит из 11 аминокислот, начиная с аминокислоты в положении 1 и заканчивая аминокислотой в положении 11. В настоящем изобретении любая ссылка на положение аминокислоты, касающееся  $\alpha$ 1-спирали или нижеописываемой консенсусной последовательности SEQ ID NO. 4, следует этой системе нумерации. Предпочтительно, чтобы немутуированная  $\alpha$ 1-спираль демонстрировала аминокислотную последовательность, как показано в Таблице 3.

**Таблица 3.** Специфические аминокислоты, входящие в состав  $\alpha$ 1-спирали белка CENH3

Положение в $\alpha$ 1-спирале	Аминокислота(ы)
1	A, F, R или S
2	A, M или S
3	S, P, T, A или C
4	F
5	I, V, M, L, S или A
6	R
7	E, T, V, L, C, Q или A
8	V или I
9	R или K
10	S, E, M, T, E, Q, G или D
11	I, V, L или 7

Наиболее предпочтительно, чтобы  $\alpha$ 1-спираль содержала консенсусную последовательность SEQ ID NO. 4, а именно:

AAPFI RLVRE I.

5            10

Как показано выше,  $\alpha$ 1-спираль содержит в своем составе специфические аминокислоты [обозначенные однобуквенным символом].

Немутированная петля1 белка CENH3 является высококонсервативной у видов растений и состоит из 13 аминокислот, начиная с аминокислоты в положении 1 и заканчивая аминокислотой в положении 13. В настоящем изобретении любая ссылка на положение аминокислоты, касающееся петли1 или нижеописываемой консенсусной последовательности SEQ ID NO. 5, следует этой системе нумерации. Предпочтительно, чтобы немутированная петля1 демонстрировала аминокислотную последовательность, как показано в Таблице 4.

**Таблица 4.** Специфические аминокислоты, входящие в состав петли1 белка CENH3

Положение в петле1	Аминокислота(ы)
1	T, S или A
2	H, Q, N, A, Y, F, G, D или E
3	M, Q, I, F, Y, A, E, N, R, L, H или G
4	L, F, V, I или Y
5	A, T, S, C или M
6	P, N, D, R, A, T, F, R, H, S или K
7	X
8	Q, Y, D, K, R, E, G, S, P, H, N или A
9	I, V или P
10	N, G, T, E или S
11	R или P
12	W или Y
13	T, Q или S

Наиболее предпочтительно, чтобы петля1 содержала консенсусную последовательность SEQ ID NO. 5, а именно:

TNFLA PXEVT RWT.

5        10        13

Как показано выше, петля1 содержит в своем составе неспецифические [обозначенные символом X] и специфические аминокислоты [обозначенные однобуквенным символом].

Немутированная  $\alpha$ 2-спираль белка CENH3 является высококонсервативной у видов растений и состоит из 29 аминокислот,

начиная с аминокислоты в положении 1 и заканчивая аминокислотой в положении 29. В настоящем изобретении любая ссылка на положение аминокислоты, касающееся  $\alpha$ 2-спиралей или нижеописываемой консенсусной последовательности SEQ ID NO. 6, следует этой системе нумерации. Предпочтительно, чтобы немутированная  $\alpha$ 2-спираль демонстрировала аминокислотную последовательность, как показано в Таблице 5.

**Таблица 5.** Специфические аминокислоты, входящие в состав  $\alpha$ 2-спиралей белка CENH3

Положение в $\alpha$ 2-спирале	Аминокислота(ы)
1	A, P, V или L
2	E, D, Q, H или L
3	A
4	L или V
5	V, L, M, I, R, Y или T
6	S или A
7	I или L
8	Q
9	E
10	A или S
11	A или T
12	E
13	D, N, F, I или Y
14	Y, F или H
15	L, I или V
16	V или I
17	G, R, E, H, N, T, E, D или Q
18	L, M или I
19	F, M или L
20	S, E, D или G
21	D, M, V, N, E, A, R или K
22	S, G, A или T
23	M, W, N или H
24	L или H
25	C или L
26	A или T
27	L или I
28	H
29	A или S

Наиболее предпочтительно, чтобы  $\alpha_2$ -спираль содержала консенсусную последовательность SEQ ID NO. 6, а именно:

AEALL ALQEA AEDFL VHLFE DAMLC AIHA.

5        10        15        20        25        29

Как показано выше,  $\alpha_2$ -спираль содержит в своем составе специфические аминокислоты [обозначенные однобуквенным символом].

Немутированная петля2 белка CENH3 является высококонсервативной у видов растений и состоит из 7 аминокислот, начиная с аминокислоты в положении 1 и заканчивая аминокислотой в положении 7. В настоящем изобретении любая ссылка на положение аминокислоты, касающееся петли2 или нижеописываемой консенсусной последовательности SEQ ID NO. 7, следует этой системе нумерации. Предпочтительно, чтобы немутированная петля2 демонстрировала аминокислотную последовательность, как показано в Таблице 6.

**Таблица 6.** Специфические аминокислоты, входящие в состав петли2 белка CENH3

Положение в петле2	Аминокислота(ы)
1	R, K или H
2	R
3	V или I
4	T
5	L, I или V
6	M или L
7	R, K, Q, L или T

Наиболее предпочтительно, чтобы петля2 содержала консенсусную последовательность SEQ ID NO. 7, а именно:

KRVTL MK.

5        7

Как показано выше, петля2 содержит в своем составе специфические аминокислоты [обозначенные однобуквенным символом].

Немутированная  $\alpha_3$ -спираль белка CENH3 является высококонсервативной у видов растений и состоит из 10 аминокислот, начиная с аминокислоты в положении 1 и заканчивая аминокислотой в

положении 10. В настоящем изобретении любая ссылка на положение аминокислоты, касающееся  $\alpha$ 3-спирали или нижеописываемой консенсусной последовательности SEQ ID NO. 8, следует этой системе нумерации. Предпочтительно, чтобы немутированная  $\alpha$ 3-спираль демонстрировала аминокислотную последовательность, как показано в Таблице 7.

**Таблица 7.** Специфические аминокислоты, входящие в состав  $\alpha$ 3-спирали белка CENH3

Положение в $\alpha$ 3-спирали	Аминокислота(ы)
1	K или R
2	D
3	F, L, I, M или W
4	E, Q или R
5	L
6	A или T
7	R
8	R
9	L или I
10	G, R или T

Наиболее предпочтительно, чтобы  $\alpha$ 2-спираль содержала консенсусную последовательность SEQ ID NO. 8, а именно:

KDFEL ARRLG.

5            10

Как показано выше,  $\alpha$ 3-спираль содержит в своем составе специфические аминокислоты [обозначенные однобуквенным символом].

Немутинированный С-терминальный домен белка CENH3 имеет разную длину. На основании изучения многочисленных видов растений (см. ниже) изобретатели определили его длину до 7 аминокислот. В настоящем изобретении любое положение аминокислоты, касающееся С-терминального домена или нижеописываемой консенсусной последовательности SEQ ID NO. 9, соответствует следующей системе нумерации. Предпочтительно, чтобы немутинированный С-терминальный домен демонстрировал аминокислотную последовательность, как показано в Таблице 8.

**Таблица 8.** Специфические аминокислоты, входящие в состав С-терминального домена белка CENH3

Положение в С-терминальном домене	Аминокислота(ы)
1	G, K, A, S или T
2	K, R, I или A
3	G, E или A
4	R, Q или V
5	P, G, I, Q, L, S или H
6	W, L, F или V
7	X

Наиболее предпочтительно, чтобы С-терминальный домен содержал консенсусную последовательность SEQ ID NO. 9, а именно:

GKGRP W.

5 6

Как показано выше, С-терминальный домен содержит в своем составе специфические аминокислоты [обозначенные однобуквенным символом].

Согласно еще одному предпочтительному варианту осуществления настоящего изобретения мутация, вызывающая изменение любой из неспецифических или специфических аминокислот, представленных в Таблице 1 или в SEQ ID NO. 1 или 2, либо в Таблице 2 или в SEQ ID NO. 3, либо в Таблице 3 или в SEQ ID NO. 4, либо в Таблице 4 или в SEQ ID NO. 5, либо в Таблице 5 или в SEQ ID NO. 6, либо в Таблице 6 или в SEQ ID NO. 7, либо в Таблице 7 или в SEQ ID NO. 8, либо в Таблице 8 или в SEQ ID NO. 9, предпочтительно замену или делецию аминокислоты (аминокислот), может приводить к получению желаемого растения, обладающего способностью образовывать гаплоидное потомство.

Неспецифическая аминокислота, указанная в Таблице 1 или в SEQ ID NO. 1 или 2, либо в Таблице 2 или в SEQ ID NO. 3, либо в Таблице 3 или в SEQ ID NO. 4, либо в Таблице 4 или в SEQ ID NO. 5, либо в Таблице 5 или в SEQ ID NO. 6, либо в Таблице 6 или в SEQ ID NO. 7, либо в Таблице 7 или в SEQ ID NO. 8, либо в Таблице 8 или в SEQ ID NO. 9, представляет собой аминокислоту, которая хотя и будучи специфической в группе определенного вида растений, тем не менее в отдельном конкретном роде растения или отдельном конкретном виде растения не является консервативной у большинства видов растений. Таким образом, неспецифическая аминокислота, указанная в SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8 или 9 или в Таблице 1, Таблице 2, Таблице 3, Таблице 4, Таблице 5, Таблице

6, Таблице 7 или Таблице 8 принадлежит к группе определенного вида растений, определенного рода растений или кциальному конкретному виду растения - хорошо определенная специфическая аминокислота, которая, по-видимому, не обнаруживается в том же месте у другого вида растения. Аминокислотная замена неспецифической аминокислоты, показанной в SEQ ID NO. 1 или в Таблице 1, означает, что в растении, а именно в определенном виде растения, специфическая, но не консервативная, аминокислота заменяется аминокислотой, отличной от аминокислоты, встречающейся в природе в том же месте данной группы определенного вида растений, определенного рода растений или в отдельном конкретном виде растения в эндогенно закодированном нативном белке CENH3 такого вида растений. Кроме того, неспецифическая аминокислота, а также специфическая аминокислота, может играть важную роль в процессах, касающихся фолдинга или стабильности белков. Изменение такой аминокислоты может приводить к мутации CENH3, заключающейся в утрате стабильности или нарушении правильного фолдинга.

Специфические аминокислоты, приведенные в Таблице 1, Таблице 2, Таблице 3, Таблице 4, Таблице 5, Таблице 6, Таблице 7 или Таблице 8, и, в частности, специфические аминокислоты SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8 или 9 представляют собой аминокислоты, которые встречаются у большинства видов растений, предпочтительно таких, что перечислены ниже, и которые поэтому весьма консервативны.

В предпочтительном варианте осуществления изобретения консенсусная последовательность SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8 или 9 собрана из последовательностей белковых сегментов видов растений, выбираемых из группы, включающей *Hordeum vulgare*, *Hordeum bulbosum*, *Sorghum bicolor*, *Saccharum officinarum*, *Zea mays*, *Setaria italica*, *Oryza minuta*, *Oriza sativa*, *Oriza australiensis*, *Oryza alta*, *Triticum aestivum*, *Secale cereale*, *Malus domestica*, *Brachypodium distachyon*, *Hordeum marinum*, *Aegilops tauschii*, *Daucus glochidiatus*, *Beta vulgaris*, *Daucus pusillus*, *Daucus muricatus*, *Daucus carota*, *Eucalyptus grandis*, *Nicotiana sylvestris*, *Nicotiana tomentosiformis*, *Nicotiana tabacum*, *Solanum lycopersicum*, *Solanum tuberosum*, *Coffea canephora*, *Vitis vinifera*, *Erythrante guttata*, *Genlisea aurea*, *Cucumis sativus*, *Morus notabilis*, *Arabidopsis arenosa*, *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Crucihamalaya himalaica*, *Crucihamalaya wallichii*, *Cardamine flexuosa*, *Lepidium virginicum*, *Capsella bursa pastoris*, *Olmariabidopsis pumila*, *Arabis hirsute*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Raphanus sativus*, *Brassica juncea*, *Brassica nigra*, *Eruca vesicaria subsp. sativa*, *Citrus sinensis*,

*Jatropha curcas*, *Populus trichocarpa*, *Medicago truncatula*, *Cicer yamashitae*, *Cicer bijugum*, *Cicer arietinum*, *Cicer reticulatum*, *Cicer judaicum*, *Cajanus cajanifolius*, *Cajanus scarabaeoides*, *Phaseolus vulgaris*, *Glycine max*, *Astragalus sinicus*, *Lotus japonicas*, *Torenia fournieri*, *Allium cepa*, *Allium fistulosum*, *Allium sativum* и *Allium tuberosum*.

В более предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты, приведенной в Таблице 1, Таблице 2, Таблице 3, Таблице 4, Таблице 5, Таблице 6, Таблице 7 или Таблице 8. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот, приведенных в Таблице 1, Таблице 2, Таблице 3, Таблице 4, Таблице 5, Таблице 6, Таблице 7 или Таблице 8, т.е. тех аминокислот, которые являются консервативными и указаны в Таблице 1, Таблице 2, Таблице 3, Таблице 4, Таблице 5, Таблице 6, Таблице 7 или Таблице 8.

Замена специфической аминокислоты, приведенной в Таблице 1, означает замену аминокислоты, выбираемой из группы, включающей:

- a) метионин в положении 1 части А,
- b) аланин в положении 2 части А,
- c) аргинин в положении 3 части А,
- d) треонин, валин, изолейцин или аланин в положении 4 части А,
- e) лизин или аргинин в положении 5 части А,
- f) гистидин, треонин, глутамин или лизин в положении 6 части А,
- g) валин, аланин, пролин, глицин, аспарагин, пролин, аргинин, серин или гистидин в положении 9 части А,
- h) треонин, аргинин, серин, лейцин, лизин, гистидин, аспарагин, аланин или пролин в положении 10 части А,
- i) аргинин, лизин, аланин, аспарагин или треонин в положении 11 части А,
- j) серин, аланин, треонин, лейцин, лизин, аргинин, аспарагиновую кислоту, аспарагин или глутаминовую кислоту в положении 12 части А,
- k) глутамин, треонин, аргинин, аланин, пролин, серин, глицин, аспарагин, валин, лизин или аргинин в положении 13 части А,

- l) пролин, треонин, аспарагиновую кислоту, глутаминовую кислоту, глутамин, серин, аспарагин, глицин, аланин, лизин, аргинин в положении 14 части А, а также
- m) аргинин, аспарагин, гистидин, валин, глицин, лизин, серин, аланин, треонин, глутаминовую кислоту, пролин в положении 15 части А;
- n) аргинин, аспарагиновую кислоту, лизин, валин, глицин, пролин, серин, глутамин, треонин или аланин в положении 1 части В,
- o) глицин, аланин, серин, лизин, аргинин, валин, треонин, пролин или глутамин в положении 2 части В,
- p) серин, треонин, лизин, валин, аргинин, глутамин, аланин, глутаминовую кислоту, глицин, пролин или аспарагиновую кислоту в положении 3 части В,
- q) глутамин, пролин, аспарагин, треонин, глутаминовую кислоту, лизин, глицин, серин, аргинин, аланин или аспарагиновую кислоту в положении 4 части В,
- r) лизин, глутамин, пролин, глицин, аспарагин, треонин, гистидин или аргинин в положении 5 части В,
- s) лизин, аргинин, глутамин или гистидин в положении 7 части В,
- t) лизин, глутамин или аргинин в положении 8 части В,
- u) серин, аланин, треонин, лизин, пролин или аргинин в положении 9 части В,
- v) тирозин, фенилаланин, гистидин, треонин, лизин, аргинин, фенилаланин или глутамин в положении 10 части В,
- w) аргинин в положении 11 части В,
- x) тирозин, аргинин, триптофан, фенилаланин, лейцин, аспарагин или серин в положении 12 части В,
- y) аргинин или лизин в положении 13 части В, а также
- z) пролин, аланин или серин в положении 14 части В.

Замена специфической аминокислоты, приведенной в Таблице 2, означает замену аминокислоты, выбираемой из группы, включающей:

- a) глицин в положении 1,

- b) треонин в положении 2,
- c) валин в положении 3,
- d) аланин в положении 4,
- e) лейцин в положении 5,
- f) лизин, триптофан или аргинин в положении 6,
- g) глутаминовую кислоту или глутамин в положении 7,
- h) изолейцин в положении 8,
- i) аргинин в положении 9,
- j) фенилаланин, тирозин или лейцин в положении 11,
- k) глутамин или аргинин в положении 12,
- l) лизин в положении 13,
- m) глутамин, серин или треонин в положении 14, а также
- n) треонин, фенилаланин, триптофан, валин, цистеин или аланин в положении 15.

Замена специфической аминокислоты, приведенной в Таблице 3, означает замену аминокислоты, выбираемой из группы, включающей:

- a) аланин, фенилаланин, аргинин или серин в положении 1,
- b) аланин, метионин или серин в положении 2,
- c) серин, пролин, треонин, аланин или цистеин в положении 3,
- d) фенилаланин в положении 4,
- e) изолейцин, валин, метионин, лейцин, серин или аланин в положении 5,
- f) аргинин в положении 6,
- g) глутаминовую кислоту, треонин, валин, лейцин, цистеин, глутамин или аланин в положении 7,
- h) валин или изолейцин в положении 8,
- i) аргинин или лизин в положении 9,

- j) серин, глутаминовую кислоту, метионин, треонин, глутаминовую кислоту, глутамин, глицин или аспарагиновую кислоту в положении 10, а также
- k) изолейцин, валин, лейцин или треонин в положении 11.

Замена специфической аминокислоты, приведенной в Таблице 4, означает замену аминокислоты, выбираемой из группы, включающей:

- a) треонин, серин или аланин в положении 1,
- b) гистидин, глутамин, аспарагин, аланин, тирозин, фенилаланин, глицин, аспарагиновую кислоту или глутаминовую кислоту в положении 2,
- c) метионин, глутамин, изолейцин, фенилаланин, тирозин, аланин, глутаминовую кислоту, аспарагин, аргинин, лейцин, гистидин или глицин в положении 3,
- d) лейцин, фенилаланин, валин, изолейцин или тирозин в положении 4,
- e) аланин, треонин, серин, цистеин или метионин в положении 5,
- f) пролин, аспарагин, аспарагиновую кислоту, аргинин, аланин, треонин, фенилаланин, аргинин, гистидин, серин или лизин в положении 6,
- g) глутамин, тирозин, аспарагиновую кислоту, лизин, аргинин, глутаминовую кислоту, глицин, серин, пролин, гистидин, аспарагин или аланин в положении 8,
- h) изолейцин, валин или пролин в положении 9,
- i) аспарагин, глицин, треонин, глутаминовую кислоту или серин в положении 10,
- j) аргинин или пролин в положении 11,
- k) триптофан, изолейцин или тирозин в положении 12, а также
- l) треонин, глутамин или серин в положении 13.

Замена специфической аминокислоты, приведенной в Таблице 5, означает замену аминокислоты, выбираемой из группы, включающей:

- a) аланин, пролин, валин или лейцин в положении 1,
- b) глутаминовую кислоту, аспарагиновую кислоту, глутамин, гистидин или лейцин в положении 2,

- c) аланин в положении 3,
- d) лейцин или валин в положении 4,
- e) валин, лейцин, метионин, изолейцин, аргинин, тирозин или треонин в положении 5,
- f) серин или аланин в положении 6,
- g) изолейцин или лейцин в положении 7,
- h) глутамин в положении 8,
- i) глутаминовую кислоту в положении 9,
- j) аланин или серин в положении 10,
- k) аланин или треонин в положении 11,
- l) глутаминовую кислоту в положении 12,
- m) аспарагиновую кислоту, аспарагин, фенилаланин, изолейцин или тирозин в положении 13,
- n) тирозин, фенилаланин или гистидин в положении 14,
- o) лейцин, изолейцин или валин в положении 15,
- p) валин или изолейцин в положении 16,
- q) глицин, аргинин, глутаминовую кислоту, гистидин, аспарагин, треонин, глутаминовую кислоту, аспарагиновую кислоту или глутамин в положении 17,
- r) лейцин, метионин или изолейцин в положении 18,
- s) фенилаланин, метионин или лейцин в положении 19,
- t) серин, глутаминовую кислоту, аспарагиновую кислоту или глицин в положении 20,
- u) аспарагиновую кислоту, метионин, валин, аспарагин, глутаминовую кислоту, аланин, аргинин, лизин в положении 21,
- v) серин, глицин, аланин или треонин в положении 22,
- w) метионин, триптофан, аспарагин или гистидин в положении 23,
- x) лейцин или гистидин в положении 24,

- у) цистеин или лейцин в положении 25,
- з) аланин или треонин в положении 26,
- аа) лейцин или изолейцин в положении 27,
- бб) гистидин в положении 28, а также
- сс) аланин или серин в положении 29.

Замена специфической аминокислоты, приведенной в Таблице 5, означает замену аминокислоты, выбираемой из группы, включающей:

- а) аргинин, лизин или гистидин в положении 1,
- б) аргинин в положении 2,
- с) валин или изолейцин в положении 3,
- д) треонин в положении 4,
- е) лейцин, изолейцин или валин в положении 5,
- ф) метионин или лейцин в положении 6, а также
- г) аргинин, лизин, глутамин, лейцин или треонин в положении 7.

Замена специфической аминокислоты, приведенной в Таблице 7, означает замену аминокислоты, выбираемой из группы, включающей:

- а) лизин или аргинин в положении 1,
- б) аспарагиновую кислоту в положении 2,
- с) фенилаланин, лейцин, изолейцин метионин иди триптофан в положении 3,
- д) глутаминовую кислоту, глутамин или аргинин в положении 4,
- е) лейцин в положении 5,
- ф) аланин или треонин в положении 6,
- г) аргинин в положении 7,
- х) аргинин в положении 8,
- и) лейцин или изолейцин в положении 9, а также
- ж) глицин, аргинин или треонин в положении 10.

Замена специфической аминокислоты, приведенной в Таблице 8, означает замену аминокислоты, выбираемой из группы, включающей:

- a) глицин, лизин, аланин, серин или треонин в положении 1,
- b) лизин, аргинин, изолейцин или аланин в положении 2,
- c) глицин, глутаминовую кислоту или аланин в положении 3,
- d) аргинин, глутамин или валин в положении 4,
- e) пролин, глицин, изолейцин, глутамин, лейцин, серин или гистидин в положении 5, а также
- f) триптофан, лейцин, фенилаланин или валин аланин в положении 6.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 1. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 1, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 1. Замена специфической аминокислоты последовательности SEQ ID NO. 1 означает замену аминокислоты, выбираемой из группы, включающей:

- a) метионин в положении 1,
- b) аланин в положении 2,
- c) аргинин в положении 3,
- d) треонин в положении 4,
- e) лизин в положении 5,
- f) гистидин в положении 6,
- g) аланин в положении 9,
- h) аргинин в положении 10,
- i) аргинин в положении 11,
- j) серин в положении 12,
- k) аргинин в положении 13,

l) лизин в положении 14, а также

m) аргинин в положении 15.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 2. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 2, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 2. Замена специфической аминокислоты последовательности SEQ ID NO. 2 означает замену аминокислоты, выбираемой из группы, включающей:

- a) глутамин в положении 1,
- b) серин в положении 2,
- c) глутамин в положении 3,
- d) треонин в положении 4,
- e) глутамин в положении 5,
- f) лизин в положении 7,
- g) лизин в положении 8,
- h) лизин в положении 9,
- i) гистидин в положении 10,
- j) аргинин в положении 11,
- k) тирозин в положении 12,
- l) аргинин в положении 13, а также
- m) пролин в положении 14.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 3. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 3, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной

последовательности SEQ ID NO. 3. Замена специфической аминокислоты последовательности SEQ ID NO. 3 означает замену аминокислоты, выбираемой из группы, включающей:

- a) глицин в положении 1,
- b) треонин в положении 2,
- c) валин в положении 3,
- d) аланин в положении 4,
- e) лейцин в положении 5,
- f) аргинин в положении 6,
- g) глутаминовую кислоту в положении 7,
- h) изолейцин в положении 8,
- i) аргинин в положении 9,
- j) фенилаланин в положении 11,
- k) глутамин или аргинин в положении 12,
- l) лизин в положении 13,
- m) треонин в положении 14, а также
- n) треонин в положении 15.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 4. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 4, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 4. Замена специфической аминокислоты последовательности SEQ ID NO. 4 означает замену аминокислоты, выбираемой из группы, включающей:

- a) аланин в положении 1,
- b) аланин в положении 2,
- c) пролин в положении 3,

- d) фенилаланин в положении 4,
- e) изолейцин в положении 5,
- f) аргинин в положении 6,
- g) аминоизокапроновую кислоту в положении 7,
- h) валин в положении 8,
- i) аргинин в положении 9,
- j) глутаминовую кислоту в положении 10, а также
- k) изолейцин в положении 11.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 5. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 5, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 5. Замена специфической аминокислоты последовательности SEQ ID NO. 5 означает замену аминокислоты, выбираемой из группы, включающей:

- a) треонин в положении 1,
- b) аспарагин в положении 2,
- c) фенилаланин в положении 3,
- d) лейцин в положении 4,
- e) аланин в положении 5,
- f) пролин в положении 6,
- g) глутаминовую кислоту в положении 8,
- h) валин в положении 9,
- i) треонин в положении 10,
- j) аргинин в положении 11,
- k) триптофан в положении 12, а также

l) треонин в положении 13.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 6. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 6, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 6. Замена специфической аминокислоты последовательности SEQ ID NO. 6 означает замену аминокислоты, выбираемой из группы, включающей:

- a) аланин в положении 1,
- b) глутаминовую кислоту в положении 2,
- c) аланин в положении 3,
- d) лейцин в положении 4,
- e) лейцин в положении 5,
- f) аланин в положении 6,
- g) лейцин в положении 7,
- h) глутамин в положении 8,
- i) глутаминовую кислоту в положении 9,
- j) аланин в положении 10,
- k) аланин в положении 11,
- l) глутаминовую кислоту в положении 12,
- m) аспарагиновую кислоту в положении 13,
- n) фенилаланин в положении 14,
- o) лейцин в положении 15,
- p) валин в положении 16,
- q) гистидин в положении 17,
- r) лейцин в положении 18,

- s) фенилаланин в положении 19,
- t) глутаминовую кислоту в положении 20
- u) аспарагиновую кислоту в положении 21,
- v) аланин в положении 22,
- w) метионин в положении 23,
- x) лейцин в положении 24,
- y) цистеин в положении 25,
- z) аланин в положении 26,
- aa) изолейцин в положении 27,
- bb) гистидин в положении 28, а также
- cc) аланин в положении 29.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 7. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 7, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 7. Замена специфической аминокислоты последовательности SEQ ID NO. 7 означает замену аминокислоты, выбираемой из группы, включающей:

- a) лизин в положении 1,
- b) аргинин в положении 2,
- c) валин в положении 3,
- d) треонин в положении 4,
- e) лейцин в положении 5,
- f) метионин в положении 6, а также
- g) лизин в положении 7.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты

последовательности SEQ ID NO. 8. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 8, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 8. Замена специфической аминокислоты последовательности SEQ ID NO. 8 означает замену аминокислоты, выбираемой из группы, включающей:

- a) лизин в положении 1,
- b) аспарагиновую кислоту в положении 2,
- c) фенилаланин в положении 3,
- d) глутаминовую кислоту в положении 4,
- e) лейцин в положении 5,
- f) аланин в положении 6,
- g) аргинин в положении 7,
- h) аргинин в положении 8,
- i) лейцин в положении 9, а также
- j) глицин в положении 10.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты последовательности SEQ ID NO. 9. Таким образом, растение по данному изобретению содержит по меньшей мере одну замену одной из специфических аминокислот SEQ ID NO. 9, т.е. аминокислот, которые являются высококонсервативными и названы в консенсусной последовательности SEQ ID NO. 9. Замена специфической аминокислоты последовательности SEQ ID NO. 9 означает замену аминокислоты, выбираемой из группы, включающей:

- a) глицин в положении 1,
- b) лизин в положении 2,
- c) глицин в положении 3,
- d) гуанидинон-аминовалериановую кислоту в положении 4,

е) пролин в положении 5, а также

ф) триптофан в положении 6.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в N-терминальном хвостовом домене, где аминокислотный аргинин в положении 3 последовательности SEQ ID NO. 1 заменен предпочтительно на лизин, либо аминокислотный аргинин в положении 2 последовательности SEQ ID NO. 23 заменен предпочтительно на лизин, либо аминокислотный аргинин в положении 10 последовательности SEQ ID NO. 1 заменен предпочтительно на фенилаланин, либо аминокислотный серин в положении 9 последовательности SEQ ID NO. 14 заменен предпочтительно на фенилаланин, либо аминокислотный аргинин в положении 16 последовательности SEQ ID NO. 14 заменен предпочтительно на глутамин, либо аминокислотный серин в положении 24 последовательности SEQ ID NO. 14 заменен предпочтительно на лейцин, либо аминокислотный аланин в положении 25 последовательности SEQ ID NO. 17 заменен предпочтительно на треонин, либо аминокислотная глутаминовая кислота в положении 29 последовательности SEQ ID NO. 14 заменена предпочтительно на лизин, либо аминокислотный глицин в положении 30 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагиновую кислоту, либо аминокислотный аланин в положении 33 последовательности SEQ ID NO. 14 или положении 32 последовательности SEQ ID NO. 20 заменен предпочтительно на треонин, либо аминокислотный пролин в положении 35 последовательности SEQ ID NO. 14 заменен предпочтительно на лейцин, либо аминокислотная глутаминовая кислота в положении 35 последовательности SEQ ID NO. 20 заменена предпочтительно на лизин, либо аминокислотный серин в положении 41 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагин, либо аминокислотный глицин в положении 43 последовательности SEQ ID NO. 14 заменен предпочтительно на глутаминовую кислоту, либо аминокислотный пролин в положении 50 последовательности SEQ ID NO. 14 заменен предпочтительно на серин, либо аминокислотный пролин в положении 55 последовательности SEQ ID NO. 14 заменен предпочтительно на лейцин, либо аминокислотный глицин в положении 57 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагиновую кислоту, либо аминокислотный глицин в положении 61 последовательности SEQ ID NO. 14 заменен предпочтительно на глутаминовую кислоту, либо аминокислотный аргинин в положении 65 последовательности SEQ ID NO. 14 заменен предпочтительно на глутамин,

либо аминокислотный аргинин в положении 65 последовательности SEQ ID NO. 14 заменен предпочтительно на стоп-сигнал, либо аминокислотный пролин в положении 71 последовательности SEQ ID NO. 14 заменен предпочтительно на серин, либо аминокислотная аспарагиновая кислота в положении 46 последовательности SEQ ID NO. 23 заменена предпочтительно на аспарагин или глицин, либо аминокислотный лизин в положении 7 последовательности SEQ ID NO. 2 заменен предпочтительно на серин, либо аминокислотный пролин в положении 56 последовательности SEQ ID NO. 20 заменен предпочтительно на серин, либо аминокислотный пролин в положении 14 последовательности SEQ ID NO. 2 заменен предпочтительно на валин, либо аминокислотный аланин в положении 62 последовательности SEQ ID NO. 17 заменен предпочтительно на валин.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в  $\alpha$ N-спирале, где аминокислотный треонин в положении 2 последовательности SEQ ID NO. 3 заменен предпочтительно на серин, либо аминокислотный треонин в положении 64 последовательности SEQ ID NO. 17 заменен предпочтительно на серин.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в  $\alpha$ 1-спирале, где аминокислотный аланин в положении 1 последовательности SEQ ID NO. 4 заменен предпочтительно на треонин, либо аминокислотный аланин в положении 105 последовательности SEQ ID NO. 14 заменен предпочтительно на треонин, либо аминокислотный аргинин в положении 6 последовательности SEQ ID NO. 4 заменен предпочтительно на глутамин, либо аминокислотный аргинин в положении 110 последовательности SEQ ID NO. 14 заменен предпочтительно на глутамин, либо аминокислотный валин в положении 89 последовательности SEQ ID NO. 20 заменен предпочтительно на метионин, либо аминокислотная глутаминовая кислота в положении 10 последовательности SEQ ID NO. 4 заменена предпочтительно на аспарагин, либо аминокислотный серин в положении 114 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагин.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в петле1, где аминокислотный аспарагин в положении 2 последовательности SEQ ID NO. 5 заменен предпочтительно на валин, либо

аминокислотный аланин в положении 95 последовательности SEQ ID NO. 17 заменен предпочтительно на валин, либо аминокислотный пролин в положении 6 последовательности SEQ ID NO. 5 заменен предпочтительно на серин, либо аминокислотный пролин в положении 121 последовательности SEQ ID NO. 14 заменен предпочтительно на серин, либо аминокислотный триптофан в положении 12 последовательности SEQ ID NO. 5 заменен предпочтительно на стоп-сигнал, либо аминокислотный триптофан в положении 127 последовательности SEQ ID NO. 14 заменен предпочтительно на стоп-сигнал.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в  $\alpha$ 2-спирале, где аминокислотный аланин в положении 1 последовательности SEQ ID NO. 6 заменен предпочтительно на треонин, либо аминокислотный аланин в положении 107 последовательности SEQ ID NO. 20 заменен предпочтительно на треонин, либо аминокислотный лейцин в положении 4 последовательности SEQ ID NO. 6 заменен предпочтительно на фенилаланин или глутамин, либо аминокислотный лейцин в положении 132 последовательности SEQ ID NO. 14 или в положении 106 последовательности SEQ ID NO. 23 заменен предпочтительно на фенилаланин или глутамин, либо аминокислотный лейцин в положении 7 последовательности SEQ ID NO. 6 заменен предпочтительно на пролин, либо аминокислотный лейцин в положении 109 последовательности SEQ ID NO. 23 заменен предпочтительно на пролин, либо аминокислотный глутамин в положении 8 последовательности SEQ ID NO. 6 заменен предпочтительно на стоп-сигнал или лейцин, либо аминокислотный глутамин в положении 114 последовательности SEQ ID NO. 20 или в положении 110 последовательности SEQ ID NO. 23 заменен предпочтительно на стоп-сигнал или лейцин, либо аминокислотный аланин в положении 10 последовательности SEQ ID NO. 6 заменен предпочтительно на треонин, либо аминокислотный аланин в положении 138 последовательности SEQ ID NO. 14 заменен предпочтительно на треонин, либо аминокислотный цистеин в положении 25 последовательности SEQ ID NO. 6 заменен предпочтительно на тирозин, либо аминокислотный цистеин в положении 153 последовательности SEQ ID NO. 14 заменен предпочтительно на тирозин, либо аминокислотный аланин в положении 26 последовательности SEQ ID NO. 6 заменен предпочтительно на валин, либо аминокислотный аланин в положении 154 последовательности SEQ ID NO. 14 заменен предпочтительно на валин.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в петле2, где аминокислотный аргинин в положении 2 последовательности SEQ ID NO. 7 заменен предпочтительно на гистидин, либо аминокислотный аргинин в положении 159 последовательности SEQ ID NO. 14 заменен предпочтительно на гистидин, либо аминокислотный валин в положении 3 последовательности SEQ ID NO. 7 заменен предпочтительно на изолейцин, либо аминокислотный валин в положении 160 последовательности SEQ ID NO. 14 заменен предпочтительно на изолейцин, либо аминокислотный треонин в положении 4 последовательности SEQ ID NO. 7 заменен предпочтительно на изолейцин, либо аминокислотный треонин в положении 139 последовательности SEQ ID NO. 20 заменен предпочтительно на изолейцин.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в  $\alpha$ 3-спирале, где аминокислотная аспарагиновая кислота в положении 2 последовательности SEQ ID NO. 8 заменена предпочтительно на аспарагин, либо аминокислотная аспарагиновая кислота в положении 166 последовательности SEQ ID NO. 14 заменена предпочтительно на аспарагин, либо аминокислотная глутаминовая кислота в положении 4 последовательности SEQ ID NO. 8 заменена предпочтительно на лизин, либо аминокислотная глутаминовая кислота в положении 168 последовательности SEQ ID NO. 14 заменена предпочтительно на лизин, либо аминокислотный аргинин в положении 8 последовательности SEQ ID NO. 8 заменен предпочтительно на гистидин, либо аминокислотный аргинин в положении 172 последовательности SEQ ID NO. 14 заменен предпочтительно на гистидин, либо аминокислотный лейцин в положении 9 последовательности SEQ ID NO. 8 заменен предпочтительно на фенилаланин, либо аминокислотный лейцин в положении 173 последовательности SEQ ID NO. 14 заменен предпочтительно на фенилаланин, либо аминокислотный глицин в положении 10 последовательности SEQ ID NO. 8 заменен предпочтительно на глутаминовую кислоту, либо аминокислотный глицин в положении 174 последовательности SEQ ID NO. 14 или в положении 152 последовательности SEQ ID NO. 20 заменен предпочтительно на глутаминовую кислоту.

Еще в одном наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация вызывает замену специфической аминокислоты в С-терминальном домене, где аминокислотный глицин в

положении 3 последовательности SEQ ID NO. 9 заменен предпочтительно на гистидин, либо аминокислотный аргинин в положении 155 последовательности SEQ ID NO. 20 заменен предпочтительно на гистидин, либо аминокислотный аргинин в положении 4 последовательности SEQ ID NO. 9 заменен предпочтительно на лизин, либо аминокислотный аргинин в положении 178 последовательности SEQ ID NO. 14 заменен предпочтительно на лизин, либо аминокислотный серин в положении 157 последовательности SEQ ID NO. 17 заменен предпочтительно на лейцин.

В предпочтительном альтернативном варианте осуществления изобретения по меньшей мере одна мутация локализуется в сайте сплайсинга геномной нуклеотидной последовательности, кодирующей белок CENH3, и/или по меньшей мере одна мутация создает новый сайт сплайсинга внутри экзона. Предпочтительно, чтобы растение, являющееся гетерозиготным в отношении такой мутации(ий), было жизнеспособным. Такая мутация(и) вызывает нарушение в сплайсинге (ошибка сплайсинга), что приводит к увеличению трансляционной клеточной продукции не полностью функционирующих белков CENH3, т.е. белков с нарушенной стабильностью, уменьшенной способностью связываться с ДНК, измененной геометрической формой, предпочтительно измененной вторичной или третичной структурой, либо с беспорядочным фолдингом, по сравнению полностью функционирующим белком CENH3 дикого типа.

В наиболее предпочтительном варианте осуществления изобретения по меньшей мере одна мутация приводит к ошибке сплайсинга, предпочтительно в инtronе 1 SEQ ID BO. 12, вызывающей изменение аминокислотной последовательности белка CENH3 *Brassica napus* после аминокислоты в положении 18 SEQ ID BO. 14, ошибке сплайсинга, предпочтительно в инtronе 2 SEQ ID BO. 12, вызывающей изменение аминокислотной последовательности белка CENH3 *Brassica napus* после аминокислоты в положении 33 SEQ ID BO. 14, ошибке сплайсинга, предпочтительно в экзоне 3 SEQ ID BO. 12, вызывающей изменение аминокислотной последовательности белка CENH3 *Brassica napus* после аминокислоты в положении 37 SEQ ID BO. 14, либо ошибке сплайсинга, предпочтительно в инtronе 8 SEQ ID BO. 12, вызывающей изменение аминокислотной последовательности белка CENH3 *Brassica napus* после аминокислоты в положении 163 SEQ ID BO. 14, либо по меньшей мере одна мутация приводит к ошибке сплайсинга, предпочтительно в инtronе 4 SEQ ID BO. 18, вызывающей изменение аминокислотной последовательности белка CENH3 *Zea mays* после аминокислоты в положении 89 SEQ ID BO. 20,

ошибке сплайсинга, предпочтительно в инtronе 5 SEQ ID BO. 18, вызывающей изменение аминокислотной последовательности белка CENH3 *Zea mays* после аминокислоты в положении 115 SEQ ID BO. 20, либо ошибке сплайсинга, предпочтительно в инtronе 6 SEQ ID BO. 18, вызывающей изменение аминокислотной последовательности белка CENH3 *Zea mays* после аминокислоты в положении 141 SEQ ID BO. 20, либо по меньшей мере одна мутация приводит к ошибке сплайсинга, предпочтительно в инtronе 1 SEQ ID BO. 15, вызывающей изменение аминокислотной последовательности белка CENH3 *Sorghum bicolor* после аминокислоты в положении 26 SEQ ID BO. 17.

В еще одном предпочтительном альтернативном варианте осуществления изобретения по меньшей мере одна мутация вызывает изменение аминокислотной последовательности белка CENH3, и это изменение придает биологическую активность гаплоидного индуктора, при этом изменение представляет собой инсерцию или делецию одной или нескольких аминокислот. В частности, инсерция может осуществляться путем мутагенеза транспозоном, а делеция – путем геномного инжиниринга. Инсерция и делеция могут происходить в любой нуклеотидной последовательности, кодирующей один из вышеупомянутых сегментов, в нуклеотидной последовательности интрана, либо в нуклеотидной последовательности 5'-нетранслируемой области (НТО) или 3'-НТО гена CENH3, при этом 5'-НТО локализуется вверх по течению от нуклеотидной последовательности, кодирующей N-терминальный хвостовой домен, и 3'-НТО локализуется вниз по течению от нуклеотидной последовательности, кодирующей С-терминальный домен. В любом случае инсерция или делеция приводит к изменению аминокислотной последовательности белка CENH3, и это изменение придает биологическую активность гаплоидного индуктора. Инсерция может иметь длину по меньшей мере 1 нуклеотид, по меньшей мере 2 нуклеотида, по меньшей мере 3 нуклеотида, по меньшей мере 4 нуклеотида, по меньшей мере 5 нуклеотидов, по меньшей мере 6 нуклеотидов, по меньшей мере 7 нуклеотидов, по меньшей мере 8 нуклеотидов, по меньшей мере 9 нуклеотидов, по меньшей мере 10 нуклеотидов, по меньшей мере 12 нуклеотидов, по меньшей мере 14 нуклеотидов, по меньшей мере 16 нуклеотидов, по меньшей мере 18 нуклеотидов, по меньшей мере 20 нуклеотидов, по меньшей мере 25 нуклеотидов, по меньшей мере 30 нуклеотидов, по меньшей мере 40 нуклеотидов, по меньшей мере 50 нуклеотидов, по меньшей мере 75 нуклеотидов, по меньшей мере 100 нуклеотидов, по меньшей мере 200

нуклеотидов, по меньшей мере 300 нуклеотидов, либо по меньшей мере 500 нуклеотидов.

В контексте настоящего изобретения термин “по меньшей мере одна мутация” означает предпочтительно одну мутацию, в частности только одну мутацию. В еще одном предпочтительном варианте осуществления изобретения термин “по меньшей мере одна мутация” означает две мутации, в частности только две мутации. Еще в одном предпочтительном варианте осуществления изобретения термин “по меньшей мере одна мутация” означает три мутации, в частности только три мутации. В еще одном предпочтительном варианте осуществления изобретения термин “по меньшей мере одна мутация” означает четыре мутации, в частности только четыре мутации. Еще в одном предпочтительном варианте осуществления изобретения термин “по меньшей мере одна мутация” означает пять мутаций, в частности только пять мутаций. Если мутаций несколько, они могут также происходить в различных полинуклеотидах и вызывать изменение аминокислотных последовательностей разных белков СЕННЗ, если они имеются у конкретного вида растений. Например, у *Hordeum vulgare* имеются два разных белка СЕННЗ.

В предпочтительном варианте осуществления настоящего изобретения по меньшей мере одна мутация представляет собой по меньшей мере одну мутацию, по меньшей мере две мутации, по меньшей мере три мутации, по меньшей мере четыре мутации или по меньшей мере пять мутаций.

Еще в одном предпочтительном варианте осуществления изобретения в одном сегменте белка СЕННЗ имеется замена одной аминокислоты, в частности замена только одной аминокислоты.

Еще в одном предпочтительном варианте осуществления изобретения в одном сегменте белка СЕННЗ имеется замена двух аминокислот, в частности замена только двух аминокислот.

Еще в одном предпочтительном варианте осуществления изобретения в одном сегменте белка СЕННЗ имеется замена трех аминокислот, в частности замена только трех аминокислот.

Еще в одном предпочтительном варианте осуществления изобретения в одном сегменте белка СЕННЗ имеется замена четырех аминокислот, в частности замена только четырех аминокислот.

Еще в одном предпочтительном варианте осуществления изобретения в одном сегменте белка СЕНЗ имеется замена пяти аминокислот, в частности замена только пяти аминокислот.

В предпочтительном варианте осуществления изобретения в одном сегменте белка СЕНЗ имеется замена 1, 1 или 2, 1-3, 1-4, 1-5, предпочтительно 1-6, наиболее предпочтительно 1-7 аминокислот.

В частности, настоящее изобретение касается мутаций, которые вызывают или приводят к замене аминокислоты внутри сегмента белка СЕНЗ. Таким образом, в контексте настоящего изобретения мутация предпочтительно представляет собой несинонимичную точечную мутацию или замену в последовательности ДНК, кодирующей белок СЕНЗ, в результате чего происходит изменение аминокислоты. Она также называется миссенс-мутацией. Кроме того, изменение аминокислоты или замена аминокислоты может быть консервативной, т.е. заменой аминокислоты аминокислотой со сходными физико-химическими свойствами, полуконсервативной, т.е. заменой отрицательно заряженной аминокислоты положительно заряженной аминокислотой, либо радикальной, т.е. заменой одной аминокислоты на совершенно другую.

В предпочтительном варианте осуществления растение по настоящему изобретению, обладающее биологической активностью гаплоидного индуктора, является гомозиготным в отношении по меньшей мере одной мутации. В еще одном варианте осуществления растение по настоящему изобретению, обладающее биологической активностью гаплоидного индуктора, является гетерозиготным в отношении по меньшей мере одной мутации.

Растение по настоящему изобретению обладает биологической активностью гаплоидного индуктора. Это означает, что скрещивание растения по изобретению с растением дикого типа или с растением, экспрессирующим белок СЕНЗ дикого типа, дает по меньшей мере 0,1%, 0,2%, 0,3%, 0,4%, 0,5%, 0,6%, 0,7%, 0,8%, 0,9%, предпочтительно по меньшей мере 1%, предпочтительно по меньшей мере 2%, предпочтительно по меньшей мере 3%, предпочтительно по меньшей мере 4%, предпочтительно по меньшей мере 5%, предпочтительно по меньшей мере 6%, предпочтительно по меньшей мере 7%, предпочтительно по меньшей мере 8%, предпочтительно по меньшей мере 9%, наиболее предпочтительно по меньшей мере 10%, по меньшей мере 15%, по меньшей мере 20% или более процентов гаплоидного потомства. При этом растение дикого типа

предпочтительно является растением того же вида, которое не содержит по меньшей мере одну мутацию растения по данному изобретению в соответствующем эндогенном гене CENH3, т.е. растение способно экспрессировать нативный белок CENH3, а растение, экспрессирующее CENH3 дикого типа, предпочтительно является растением того же вида, которое содержит i) полинуклеотид, включающий нуклеотидную последовательность, кодирующую белок CENH3 без по меньшей мере одной мутации растения по данному изобретению и способно экспрессировать такой нативный белок CENH3, либо ii) полинуклеотид, включающий нуклеотидную последовательность, кодирующую белок CENH3 из другого вида растения, проявляющего сравнимую функциональность нативного CENH3, например, белок CENH3 из другого вида растения может быть интродуцирован в качестве трансгена.

Таким образом, настоящее изобретение наилучшим образом обеспечивает средства и методы для создания линий-индукторов гаплоидов в обширном диапазоне настоящих двудольных (эвдикотов), двудольных и однодольных видов растений. Настоящее изобретение также позволяет осуществлять замену материнской цитоплазмы и, например, создавать растения с цитоплазматической мужской стерильностью желаемого генотипа за один этап. К преимуществам настоящего изобретения также относится то, что посредством мутагенеза или любого иного подхода, не связанного с использованием ГМО, можно вызывать мутацию единственной аминокислоты.

Таким образом, в предпочтительном варианте осуществления изобретения весь процесс гаплоидизации посредством применения линии-индуктора гаплоидов, характеризующейся точечно мутированным эндогенным геном CENH3, с заменой аминокислоты по меньшей мере в одном из положений, обеспечиваемой настоящим изобретением, не является трансгенным.

В контексте настоящего изобретения “эндогенный” ген, аллель или белок относится к нерекомбинантной последовательности в растении, поскольку эта последовательность встречается в соответствующем растении, в частности в дикорастущем растении. Термин “мутированный” относится к последовательности, измененной человеком. Примером нетрансгенной мутации, вызываемой человеком, является воздействие на растение большой дозой химических, радиологических агентов или иных мутагенов с целью отбора мутантов. В противоположность этому трансгенные мутации,

вызываемые человеком, т.е. рекомбинантные изменения или технологии генной инженерии, например, посредством TALE-нуклеаз, нуклеаз “цинковые пальцы” или системы CRISPR/Cas, включают слияния, инсерции, делеции и/или изменения ДНК или аминокислотной последовательности.

Полинуклеотидная или полипептидная последовательность является “гетерологичной или экзогенной по отношению к” организму, если она происходит из чужеродного вида, либо, если она происходит из того же вида, изменена ее исходная форма. “Рекомбинантная” означает измененная человеком, т.е. трансгенная полинуклеотидная или полипептидная последовательность. Термин “трансген” имеет то значение, в каком он используется специалистами в данной области, и относится к предпочтительно гетерологичной нуклеиновой кислоте, интродуцируемой в клетку посредством молекулярного манипулирования клеточным геномом, осуществляемого человеком, т.е. посредством молекулярной трансформации. Таким образом, “трансгенное растение” представляет собой растение, содержащее трансген, т.е. генетически модифицированное растение. Трансгенное растение может представлять собой исходное растение, в которое трансген был введен, а также его потомство, в геноме которого также содержится трансген.

Термин “нуклеотидная последовательность, кодирующая” относится к нуклеиновой кислоте, которая направляет экспрессию специфического белка, в частности белка СЕННЗ или его частей. Нуклеотидные последовательности включают как последовательность цепи ДНК, транскрибуемую в РНК, так и последовательность РНК, транслируемую в белок. Нуклеотидные последовательности включают полноразмерные последовательности нуклеиновой кислоты, а также последовательности неполной длины, получаемые из полноразмерных последовательностей.

Термин “ген” относится к кодирующей нуклеотидной последовательности и связанным регуляторным нуклеотидным последовательностям, инtronу(ам), 5' -НТО и/или 3'-НТО.

Термин “регуляторный элемент” касается последовательности, предпочтительно нуклеотидной последовательности, расположенной вверх по течению (5'), внутри, и/или вниз по течению (3') от нуклеотидной последовательности, предпочтительно кодирующей последовательности, транскрипция и экспрессия которой контролируется регуляторными элементами потенциально во взаимосвязи с биосинтетическим аппаратом клетки белка. Термины “регуляция” или “регулировать” касаются модуляции

экспрессии генов, индуцируемой элементами последовательности ДНК, располагающимися прежде всего, но не исключительно, вверх по течению (5') от начала транскрипции интересующих генов. Результатом регуляции может быть проявление экспрессии или отсутствие проявления экспрессии генов в ответ на стимуляцию, а также изменение уровня экспрессии генов.

Регуляторный элемент, в частности последовательность ДНК, такой как промотор, считается “оперативно присоединенным к” или “оперативно связанным с” последовательностью ДНК, которая кодирует РНК или белок, если обе последовательности расположены и ориентированы таким образом, что регуляторная последовательность ДНК вызывает экспрессию кодирующей последовательности ДНК.

“Промотор” представляет собой последовательность ДНК, инициирующую транскрипцию связанной последовательности ДНК, в частности последовательности, которая располагается вверх по течению (5') от начала транскрипции, вовлечена в распознавание и представляет собой РНК-полимеразу. В зависимости от конкретного участка промотор может также включать элементы, которые действуют как регуляторы экспрессии генов, такие как активаторы, энхансеры и/или репрессоры.

“3' регуляторный элемент” (или 3'-конец) относится к части гена, содержащей сегмент ДНК, за исключением 5'-последовательности, обеспечивающей инициацию транскрипции, и структурной части гена, правильно определяющей сайт терминации и содержащей сигнал полиаденилирования и любые другие регуляторные сигналы, способные осуществлять процессинг матричной РНК (мРНК) или экспрессию генов. Сигнал полиаденилирования обычно характеризуется как процесс присоединения трактов полиадениловой кислоты к 3'-концу предшественника мРНК. Сигналы полиаденилирования обычно распознаются по наличию гомологии в канонической форме 5'-АТААА-3'.

Термин “кодирующая последовательность” относится к области гена, кодирующей белок, полипептид, либо его части, за исключением регуляторных последовательностей, обеспечивающих инициацию или терминацию транскрипции.

Обычно в клетках могут находиться гены, кодирующие последовательности или регуляторные элементы и в таком случае они называются “аутологичными” или “эндогенными”. Также обычно в

клеточном содержимом они могут отсутствовать и в таком случае клетки называются “гетерологичными”, “трансгенными” или “трансгенами”.

“Гетерологичные” гены, кодирующие последовательности или регуляторные элементы могут также быть аутологичными по отношению к клеткам, однако располагаться в порядке и/или ориентации, либо находиться в геноме или условиях, которые обычно отсутствуют или не встречаются в клетках, в которые они перенесены.

Термин “вектор” относится к конструкции на основе рекомбинантной ДНК, которая может представлять собой плазмиду, вирус, автономно реплицирующуюся последовательность, искусственную хромосому, такую как искусственная бактериальная хромосома ВАС, фаг или иную нуклеотидную последовательность, где соединены или рекомбинированы по меньшей мере две последовательности нуклеотидов, из которых по меньшей мере одна – это молекула нукleinовой кислоты по данному изобретению. Вектор может быть линейным или круговым. Вектор может состоять из одно- или двухцепочечной ДНК или РНК.

Термин “экспрессия” относится к транскрипции и/или трансляции эндогенного гена или трансгена в растениях.

Термины “трансформация”, “трансформирование” и “перенос” касаются методов переноса молекул нукleinовой кислоты, в частности ДНК, в клетки, включая, но не ограничиваясь, биолистические подходы, такие как бомбардировка частицами, микроинъекцию, пермеабилизацию клеточной мембранны под действием различных физических, например электропорации, или химических факторов, например, обработки полиэтиленгликолем (ПЭГ); слияние протопластов или трансформацию, опосредованную *Agrobacterium tumefaciens* или *rhizogenes*. Что касается инъекции и электропорации ДНК в растительные клетки, то к используемым плазмидам не предъявляются какие-либо определенные требования. Могут применяться такие плазмиды, как производные pUC. Если регенерации из трансформированных таким образом клеток подлежат целые растения, то предпочтительно использовать селективный маркер. В зависимости от применяемого метода введения желаемых генов в клетки растений могут понадобиться дополнительные последовательности ДНК; если для трансформации растительных клеток применяются Ti- или Ri-плазмиды, то по меньшей мере правая граница, однако часто правая и левая границы, Т-ДНК Ti- и Ri-плазмид должна соединяться в качестве flankирующей области с подлежащими интродукции генами. Предпочтительно, чтобы переносимые молекулы нукleinовой

кислоты стабильно интегрировались в геном или пластом растения-реципиента.

В контексте настоящего изобретения термины “биологическая активность гаплоидного индуктора”, “гаплоиндуктор” или “линия-индуктор гаплоидов” касаются растения или линии растений, обладающих способностью производить гаплоидное потомство по меньшей мере в 0,1%, по меньшей мере в 0,2%, 0,3%, 0,4%, 0,5%, 0,6%, 0,7%, 0,8%, 0,9%, предпочтительно по меньшей мере в 1%, предпочтительно по меньшей мере в 2%, предпочтительно по меньшей мере в 3%, предпочтительно по меньшей мере в 4%, предпочтительно по меньшей мере в 5%, предпочтительно по меньшей мере в 6%, предпочтительно по меньшей мере в 7%, предпочтительно по меньшей мере в 8%, предпочтительно по меньшей мере в 9%, наиболее предпочтительно по меньшей мере в 10%, наиболее предпочтительно по меньшей мере в 15%, наиболее предпочтительно по меньшей мере в 20% случаев при скрещивании с растением дикого типа или с растением, которое по меньшей мере экспрессирует белок СЕНН3 дикого типа. Так как в процессе мейоза хромосомы гаплоиндуктора элиминируются, то в результате в гаплоидном потомстве содержатся хромосомы только растения дикого типа. Однако если при скрещивании гаплоиндуктор являлся материнским родителем, то гаплоидное потомство будет содержать цитоплазму индуктора и хромосомы родителя дикого типа.

Термин “растение” согласно настоящему изобретению включает целые растения или части целого растения.

Предпочтительно, чтобы целые растения являлись семенными растениями, либо зерновой культурой. Части растения представляют собой, например, вегетативные органы/структуры побега, например, листья, стебли и клубни; корни, цветки и цветковые органы/структуры, например, прицветники, чашелистики, лепестки, тычинки, плодолистики, пыльники и семяпочки; семя, включая зародыш, эндосперм и семенную оболочку; соплодие и зрелую завязь; растительную ткань, например, проводящую ткань, основную паренхиму и тому подобное; и клетки, например, замыкающие клетки, яйцеклетки, трихомы и тому подобное; а также их потомство.

В любом случае растение по данному изобретению содержит по меньшей мере одну клетку, содержащую полинуклеотид, включающий нуклеотидную последовательность, кодирующую центромерный гистоновый H3 (СЕНН3) белок, при этом полинуклеотид содержит по меньшей мере одну

мутацию, вызывающую изменение аминокислотной последовательности белка СЕНЗ, которое придает биологическую активность гаплоидного индуктора растению, предпочтительно такому, как здесь описано более подробно. Наиболее предпочтительно, чтобы большая часть или, в частности, все растительные клетки по данному изобретению имели мутацию(и), как здесь описано.

Предпочтительно, чтобы вид растений, которые могут использоваться в методах по данному изобретению, принадлежал к настоящим двудольным, двудольным и однодольным растениям.

Термин “растение” в предпочтительном варианте осуществления изобретения относится исключительно к целому растению, т.е. растению, представляющему собой фенотип полностью развитого растения и способного к размножению, растению в ранней стадии развития, например, растительному зародышу, либо к тому и другому.

В варианте осуществления настоящего изобретения термин “растение” относится к части целого растения, в частности к растительному материалу, растительным клеткам или культурам растительных клеток.

Термин “растительная клетка” обозначает структурную и физиологическую единицу растения и состоит из протопласта и клеточной стенки. Растительная клетка может быть в форме изолированной единственной клетки, такой как замыкающая клетка (в устьице) или культивируемая клетка, либо как часть более высокоорганизованной единицы, например, такой как растительная ткань или растительный орган.

Термин “растительный материал” включает части растения, в частности растительные клетки, растительную ткань, в частности материал для размножения растений, предпочтительно листья, стебли, корни, появившиеся зародышевые корни, цветки или цветковые части, лепестки, соплодия, пыльцу, пыльцевые трубки, тычиночные нити, семязачатки, зародышевые мешки, яйцеклетки, завязи, зиготы, зародыши, зиготные зародыши как таковые, соматические зародыши, сегменты гипокотиля, апикальные меристемы, сосудистые пучки, перициклы, семена, корни, культуры клеток или тканей, либо любую иную часть или продукт растения.

Таким образом, настоящее изобретение также обеспечивает материал для размножения растений по данному изобретению. Указанный “материал для размножения растений” следует понимать как означающий любой растительный материал, который можно размножить как половым, так и

вегетативным путем в культуре *in vitro* или *in vivo*. В объеме настоящего изобретения наиболее предпочтительны протопласты, клетки, каллусы, ткани, органы, семена, зародыши, пыльца, яйцеклетки, зиготы наряду с любым иным материалом для размножения, полученным из трансгенных растений. Целью настоящего изобретения также являются части растений, например, такие как цветки, стебли, плоды, листья, корни, происходящие из мутированных растений или их ранее мутированных потомков, предпочтительно трансформированных с помощью методов по настоящему изобретению и поэтому состоящих по меньшей мере частично из мутированных клеток.

Предпочтительно растение по данному изобретению выбирают из группы, включающей ячмень обыкновенный (*Hordeum vulgare*), сорго двуцветное (*Sorghum bicolor*), рожь (*Secale cereale*), тритикале, сахарный тростник (*Saccharum officinarum*), кукурузу (*Zea mays*), могар (*Setaria italica*), рис посевной (*Oryza sativa*), *Oryza minuta*, *Oriza australiensis*, *Oryza alta*, пшеницу мягкую (*Triticum aestivum*), *Triticum durum*, *Hordeum bulbosum*, трахинию двуколосковую (*Brachypodium distachyon*), морской ячмень (*Hordeum marinum*), эгилопс Тауша (*Aegilops tauschii*), яблоню домашнюю (*Malus domestica*), *Beta vulgaris*, подсолнечник однолетний (*Helianthus annus*), австралийскую морковь (*Daucus glochidiatus*), американскуюдикую морковь (*Daucus pusillus*), *Daucus muricatus*, морковь обыкновенную (*Daucus carota*), эвкалипт большой (*Eucalyptus grandis*), *Erythranthe guttata*, *Genlisea aurea*, табак лесной (*Nicotiana sylvestris*), *Nicotiana tomentosiformis*, табак обыкновенный (*Nicotiana tabacum*), томат (*Solanum lycopersicum*), картофель (*Solanum tuberosum*), кофе конголезский (*Coffea canephora*), виноград культурный (*Vitis vinifera*), огурец обыкновенный (*Cucumis sativus*), шелковицу (*Morus notabilis*), резуховидку песчаную (*Arabidopsis arenosa*), *Arabidopsis lyrata*, резуховидку Талия (*Arabidopsis thaliana*), *Crucihamalaia himalaica*, *Crucihamalaia wallichii*, сердечник извилистый (*Cardamine flexuosa*), клоповник виргинский (*Lepidium virginicum*), пастушью сумку обыкновенную (*Capsella bursa-pastoris*), *Olmariabidopsis pumila*, резуху волосистую (*Arabis hirsuta*), рапс (*Brassica napus*), капусту огородную (*Brassica oleracea*), *Brassica rapa*, редьку посевную (*Raphanus sativus*), *Brassica juncea*, горчицу черную (*Brassica nigra*), *Eruca vesicaria sativa*, апельсин (*Citrus sinensis*), *Jatropha curcas*, *Glycine max* и тополь волосистоплодный (*Populus trichocarpa*).

Наиболее предпочтительно растение выбирают из группы, включающей ячмень обыкновенный (*Hordeum vulgare*), сорго двуцветное

(*Sorghum bicolor*), рожь (*Secale cereale*), тритикале, сахарный тростник (*Saccharum officinarium*), кукурузу (*Zea mays*), рис посевной (*Oryza sativa*), пшеницу мягкую (*Triticum aestivum*), *Triticum durum*, *Avena sativa*, *Hordeum bulbosum*, *Beta vulgaris*, подсолнечник однолетний (*Helianthus annus*), морковь обыкновенную (*Daucus carota*), табак обыкновенный (*Nicotiana tabacum*), томат (*Solanum lycopersicum*), картофель (*Solanum tuberosum*), кофе конголезский (*Coffea canephora*), виноград культурный (*Vitis vinifera*), огурец обыкновенный (*Cucumis sativus*), резуховидку Таля (*Arabidopsis thaliana*), рапс (*Brassica napus*), капусту огородную (*Brassica oleracea*), *Brassica rapa*, *Brassica juncea*, горчицу черную (*Brassica nigra*), редьку посевную (*Raphanus sativus*) и *Glycine max*.

В предпочтительном варианте осуществления растение по данному изобретению содержит нуклеотидную последовательность, кодирующую СЕННЗ, являющийся либо эндогенным геном, либо трансгеном.

В предпочтительном варианте осуществления изобретение касается растения, получаемого с использованием принципов по данному изобретению, где по меньшей мере одна замена аминокислоты интродуцируется в нуклеотидную последовательность, кодирующую СЕННЗ, нетрансгенно или трансгенно.

Предпочтителен вариант осуществления изобретения, где по меньшей мере одной мутации подвергается эндогенный ген СЕННЗ, при этом получаемое растение не является трансгенным. Предпочтительно, чтобы мутация осуществлялась посредством нетрансгенного мутагенеза, мутагенеза транспозоном, в частности химического мутагенеза, предпочтительно с применением ЭМС (этилметансульфонат)-индуцированного TILLING или целевого редактирования генома.

Таким образом, настоящее изобретение относится к растению, в котором нетрансгенное введение по меньшей мере одной мутации, вызывающее изменение аминокислотной последовательности белка СЕННЗ, которое придает биологическую активность гаплоидного индуктора, осуществляется с применением химического мутагенеза, в частности с применением методики TILLING.

Еще в одном предпочтительном варианте осуществления изобретения введение по меньшей мере одной мутации в растение осуществляется в форме трансгена. Предпочтительно, чтобы это происходило посредством трансформации вектора, содержащего полинуклеотид, включающий

нуклеотидную последовательность, кодирующую по меньшей мере сегмент СЕННЗ, содержащий по меньшей мере одно изменение аминокислотной последовательности, предпочтительно такое, как здесь описано. Методы трансформации растения и методы встраивания трансгена в геном растения хорошо известны специалисту в данной области техники.

В предпочтительном варианте осуществления изобретения обеспечивается растение, в котором трансгенное внесение изменения в аминокислотную последовательность белка СЕННЗ происходит посредством трансформации вектора, содержащего полинуклеотид, включающий нуклеотидную последовательность, кодирующую по меньшей мере сегмент белка СЕННЗ или домена САТД белка СЕННЗ, имеющего замену по меньшей мере одной аминокислотной последовательности, предпочтительно имеющего замену по меньшей мере одной аминокислотной последовательности одной из специфических аминокислот консенсусной последовательности SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8 или 9, либо приведенных в Таблице 1, Таблице 2, Таблице 3, Таблице 4, Таблице 5, Таблице 6, Таблице 7 или Таблице 8.

Предпочтительно, чтобы трансформация проводилась с применением *Agrobacterium*-опосредованного метода, методов погружения цветочных почек или бомбардировки частицами.

В предпочтительном варианте осуществления изобретения трансформация полинуклеотида, включающего нуклеотидную последовательность, кодирующую измененный белок СЕННЗ по данному изобретению, происходит в растении в форме введения трансгена, при этом один или оба аллеля эндогенного гена СЕННЗ предпочтительно инактивируются, либо нокаутируются. В еще одном предпочтительном варианте осуществления изобретения трансформация полинуклеотида, включающего нуклеотидную последовательность, кодирующую измененный белок СЕННЗ по данному изобретению, происходит в растении в форме введения трансгена, при этом имеет место сверхэкспрессия трансгена, чтобы он мог быть более компетентным, чем эндогенный белок СЕННЗ, и предпочтительным в процессе формирования кинетохорного комплекса.

Настоящее изобретение также обеспечивает растение, получаемое, в частности полученное, с использованием метода по данному изобретению и характеризующееся тем, что оно обладает биологической активностью гаплоидного индуктора.

В предпочтительном варианте осуществления настоящего изобретения метод получения растения, обладающего биологической активностью гаплоидного индуктора по данному изобретению, не является по существу биологическим методом.

Настоящее изобретение также обеспечивает метод создания растения, обладающего биологической активностью гаплоидного индуктора по данному изобретению, состоящий из следующих стадий:

- i) обработка семян растения достаточным количеством мутагенного этилметансульфоната (ЭМС) для получения растений М1,
- ii) обеспечение производства достаточного количества фертильных растений М2,
- iii) выделение геномной ДНК из растений М2 и
- iv) отбор индивидуумов, обладающих по меньшей мере мутацией, вызывающей изменение аминокислотной последовательности СЕНЗ.

Настоящее изобретение также касается предпочтительного варианта метода создания растения, обладающего биологической активностью гаплоидного индуктора по данному изобретению, состоящего из следующих стадий:

хх) обеспечение вектора, содержащего полинуклеотид, включающий нуклеотидную последовательность, кодирующую по меньшей мере один сегмент аминокислотной последовательности белка СЕНЗ, при этом полинуклеотид содержит по меньшей мере одну мутацию, вызывающую аминокислотной последовательности белка СЕНЗ,

уу) трансформация растительной клетки вектором, при этом растительная клетка предпочтительно содержит один или оба эндогенных аллеля гена СЕНЗ, которые инактивированы или нокаутированы, и

zz) регенерация растения, обладающего биологической активностью гаплоидного индуктора, из растительной клетки.

Настоящее изобретение касается еще одного предпочтительного варианта метода создания растения, обладающего биологической активностью гаплоидного индуктора по данному изобретению, состоящего из следующих стадий:

уу) трансформация растительной клетки посредством полинуклеотида, включающего нуклеотидную последовательность, кодирующую по меньшей мере один сегмент белка СЕННЗ, при этом полинуклеотид содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности белка СЕННЗ, либо вектора, содержащего полинуклеотид, включающий нуклеотидную последовательность, кодирующую по меньшей мере один сегмент белка СЕННЗ, при этом полинуклеотид содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности белка СЕННЗ, и

zz) регенерация растения, обладающего биологической активностью гаплоидного индуктора, из растительной клетки.

В частности, настоящее изобретение относится к гаплоидному растению, получаемому, в частности полученному, путем:

а) скрещивания растения, обладающего биологической активностью гаплоидного индуктора по данному изобретению с растением, экспрессирующим белок СЕННЗ дикого типа, и, как вариант,

б) идентификации гаплоидного потомства, создаваемого на стадии скрещивания.

Предпочтительно, чтобы идентифицированное гаплоидное растение можно было переводить в удвоенный гаплоид, предпочтительно в результате обработки колхицином, что также является частью настоящего изобретения. Таким образом, настоящее изобретение также касается удвоенного гаплоида, получаемого, в частности полученного, путем перевода гаплоидного растения по данному изобретению в удвоенный гаплоид в результате обработки колхицином или спонтанного удвоения хромосом.

Настоящее изобретение также обеспечивает метод создания гаплоидного растения, состоящий из следующих стадий:

а) скрещивание растения, обладающего биологической активностью гаплоидного индуктора по данному изобретению с растением, экспрессирующим белок СЕННЗ дикого типа, и

б) идентификация гаплоидного потомства, создаваемого на стадии скрещивания.

На следующей стадии с) отобранное гаплоидное растение предпочтительно переводят в удвоенный гаплоид, предпочтительно в

результате обработки колхицином. Таким образом, настоящее изобретение также касается метода создания удвоенного гаплоида.

В предпочтительном варианте осуществления настоящего изобретения обеспечиваемый способ не является по существу биологическим способом .

В частности, способы, о которых идет речь, не основываются только на, в частности не заключаются в, таких естественных факторах, как скрещивание или селекция, а фактически на технических принципах получения специфически мутированной нуклеотидной последовательности, осуществляемого с участием человека. В соответствии с настоящим изобретением в нуклеотидную последовательность и растение по изобретению вводится специфический структурный признак, а именно мутация, которая не вызвана посредством или не связана с каким-нибудь естественным фактором, таким как скрещивание или селекция.

Как частный случай осуществления настоящего изобретения обеспечивается способ, включающий стадию скрещивания, при этом указанная стадия скрещивания не дает (что обычно бывает при скрещивании) гетерозиготное потомство, а по существу гомозиготное потомство. Кроме того, гаплоидия у потомства не является результатом смешения генов растений, использованных для полового скрещивания. Более того, заявляемый в настоящем изобретении процесс создания удвоенного гаплоида не встречается в природе.

Настоящее изобретение также обеспечивает способ облегчения цитоплазматического обмена, состоящий из следующих стадий:

х) скрещивание растения по данному изобретению в качестве материнского родителя с растением, экспрессирующим белок СЕНЗ дикого типа в качестве отцовского родителя, и

у) получение потомства гаплоидного растения, содержащего хромосомы отцовского родителя и цитоплазму материнского родителя.

В предпочтительном варианте осуществления настоящего изобретения обеспечиваемый способ не является по существу биологическим способом. Указанный способ не является по существу биологическим способом по тем же причинам, которые были указаны выше, в частности потому, что он не основывается на таких естественных факторах, как скрещивание и селекция, а, в основном, в значительной степени на технических принципах получения конкретной мутации в нуклеотидной последовательности и растении по

данному изобретению. Кроме того, гаплоидия у потомства не является результатом смешения генов растений, использованных для полового скрещивания.

Способ может с успехом применяться для формирования цитоплазмической мужской стерильности (ЦМС). ЦМС вызывается внеядерным геном (митохондрий или хлоропластов) и наследуется по материнской линии. Таким образом, растение по данному изобретению должно проявлять ЦМС и быть материнским родителем скрещивания. Подобным образом ЦМС можно вводить в партнера по скрещиванию, предпочтительно представляющего собой элитную линию культуры.

В предпочтительном варианте осуществления растение по данному изобретению можно также использовать в способ восстановления мужской fertильности путем введения нормальной цитоплазмы в партнера по скрещиванию, проявляющего ЦМС. В результате такого скрещивания хромосомы растения, проявляющего ЦМС, вводятся в нормальную цитоплазму гаплоидного индуктора по данному изобретению, который не проявляет ЦМС. Однако при этом образование пыльцы у растения, проявляющего ЦМС, необходимо индуцировать посредством температуры, освещенности, продолжительности дня и т.п.

Не ограничиваясь только теорией, возможная модель того, как способы по данному изобретению работают и, в частности, способ элиминации хромосомы одного из родителей в индукторе *CENY3 x дикий тип CENH3* межвидовых гибридных зародышей, выглядит следующим образом: (A) Предположительно полученные из гаплоидного индуктора яйцеклетки содержат либо меньшее количество CENH3, либо по сравнению с диким типом уменьшенное количество неизвестной “сигнатуры, требуемой для CENH3-трансгенерации”. Уменьшенное количество материнского CENH3 менее возможно, так как исследования, проведенные с репортерным геном CENH3-GFP, показывают, что ядра спермиев, а не яйцеклетки, маркируются CENH3. Однако все еще возможно, что остаточные материнские CENH3, генерирующие “центромерный импринтинг”, передаются потомству. (B) В течение нескольких часов после оплодотворения отцовский CENH3 дикого типа также активно удаляется из зиготного ядра и (C) центромерная повторная загрузка CENH3-GFP в зиготу происходит на стадии 16-ядерного развития эндосперма у *A. thaliana*. (D) В зародышах, подвергающихся гаплоидизации, центромерная повторная загрузка материнских хромосом нарушается или замедляется, что приводит к расхождению хромосом

вследствие неактивности центромеры в анафазе. Как следствие, микроядерные хромосомы гаплоидного индуктора деградируют и (E) развивается гаплоидный зародыш. Гаплоидные зародыши содержат хромосомы, полученные от отцовской формы, на фоне цитоплазмы, полученной от материнской формы.

Настоящее изобретение также относится к полинуклеотиду, включающему нуклеотидную последовательность, кодирующую по меньшей мере один сегмент аминокислотной последовательности белка СЕННЗ или белок СЕННЗ, при этом полинуклеотид содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности белка СЕННЗ.

Настоящее изобретение также относится к вектору, в частности вирусному вектору, конструкции или плазмиде, которые содержат указанный полинуклеотид, а также, если имеются, связанные последовательности, предпочтительно такие, как здесь описано.

В более предпочтительном варианте осуществления настоящего изобретения полинуклеотид, включающий нуклеотидную последовательность, кодирующую сегмент белка СЕННЗ, предпочтительно содержит по меньшей мере полную кодирующую область СЕННЗ, в частности ген СЕННЗ.

В еще одном предпочтительном варианте осуществления настоящего изобретения полинуклеотид или кодирующая последовательность СЕННЗ может связываться с регуляторными элементами, такими как 5'- и/или 3'-регуляторные элементы, наиболее предпочтительно с промотором, предпочтительно конститтивным или индуцируемым промотором.

Растительная клетка, содержащая указанный полинуклеотид или вектор в качестве трансгена, является клеткой по данному изобретению.

В контексте настоящего изобретения под используемым здесь термином “содержащий” следует понимать “включающий” или “имеющий”, что означает, что кроме явно указанного элемента возможно присутствуют и другие элементы.

В предпочтительном варианте осуществления настоящего изобретения под используемым здесь термином “включающий” следует также понимать “состоящий из”, что исключает присутствие других элементов, кроме явно указанного элемента.

В еще одном предпочтительном варианте осуществления настоящего изобретения под используемым здесь термином “включающий” следует также понимать “состоящий по существу из”, что исключает присутствие других элементов, подчеркивающих значимость раскрываемого принципа, кроме явно указанного элемента.

Другие предпочтительные варианты осуществления настоящего изобретения являются предметом зависимых пунктов формулы изобретения.

Изобретение далее будет раскрыто более подробно посредством не ограничивающих примеров и рисунка.

Протокол последовательностей включает:

SEQ ID NO. 1: консенсусная аминокислотная последовательность N-терминального хвостового домена CENH3 (часть А),

SEQ ID NO. 2: консенсусная аминокислотная последовательность N-терминального хвостового домена CENH3 (часть В),

SEQ ID NO. 3: консенсусная аминокислотная последовательность  $\alpha$ N-спирали CENH3,

SEQ ID NO. 4: консенсусная аминокислотная последовательность  $\alpha$ 1-спирали CENH3,

SEQ ID NO. 5: консенсусная аминокислотная последовательность петли1 CENH3,

SEQ ID NO. 6: консенсусная аминокислотная последовательность  $\alpha$ 2-спирали CENH3,

SEQ ID NO. 7: консенсусная аминокислотная последовательность петли2 CENH3,

SEQ ID NO. 8: консенсусная аминокислотная последовательность  $\alpha$ 3-спирали CENH3,

SEQ ID NO. 9: консенсусная аминокислотная последовательность C-терминального домена CENH3,

SEQ ID NO. 10: нуклеотидная последовательность кодирующей последовательности (кДНК) CENH3 дикого типа *A. thaliana*,

SEQ ID NO. 11: аминокислотная последовательность СЕННЗ дикого типа *A. thaliana*,

SEQ ID NO. 12: нуклеотидная последовательность геномной последовательности (геномной ДНК) СЕННЗ дикого типа *B. napus*,

SEQ ID NO. 13: нуклеотидная последовательность кодирующей последовательности (кДНК) СЕННЗ дикого типа *B. napus*,

SEQ ID NO. 14: аминокислотная последовательность СЕННЗ дикого типа *B. napus*,

SEQ ID NO. 15: нуклеотидная последовательность геномной последовательности (геномной ДНК) СЕННЗ дикого типа *S. bicolor*,

SEQ ID NO. 16: нуклеотидная последовательность кодирующей последовательности (кДНК) СЕННЗ дикого типа *S. bicolor*,

SEQ ID NO. 17: аминокислотная последовательность СЕННЗ дикого типа *S. bicolor*,

SEQ ID NO. 18: нуклеотидная последовательность геномной последовательности (геномной ДНК) СЕННЗ дикого типа *Z. mays*,

SEQ ID NO. 19: нуклеотидная последовательность кодирующей последовательности (кДНК) СЕННЗ дикого типа *Z. mays*,

SEQ ID NO. 20: аминокислотная последовательность СЕННЗ дикого типа *Z. mays*,

SEQ ID NO. 21: нуклеотидная последовательность геномной последовательности (геномной ДНК) СЕННЗ дикого типа *B. vulgaris*,

SEQ ID NO. 22: нуклеотидная последовательность кодирующей последовательности (кДНК) СЕННЗ дикого типа *B. vulgaris*,

SEQ ID NO. 23: аминокислотная последовательность СЕННЗ дикого типа *B. vulgaris*, а также

SEQ ID NO. 24: нуклеотидная последовательность геномной последовательности (геномной ДНК) Ми-мутированного СЕННЗ *Z. mays*.

На рисунке представлено выравнивание аминокислотных последовательностей *Arabidopsis thaliana* (первый ряд), *Beta vulgaris* (второй ряд), *Brassica napus* (третий ряд), *Zea mays* (четвертый ряд), *Sorghum bicolor*

(пятый ряд), а также диаграмма, показывающая уровень консервативности этих пяти видов растений.

## **Примеры**

### Идентификация мутантов CENH3

Для идентификации мутаций внутри гена CENH3, вызывающих изменение аминокислотной последовательности транслируемого CENH3, где изменение способно придавать биологическую активность гаплоидного индуктора растению, исследовали все сегменты гена CENH3 в отношении подходящих мутаций, несмотря на то, что Ravi и Chan (2010) подчеркивали особую важность только N-терминального домена. Исследования мутантов в других сегментах, таких как  $\alpha$ 2-спираль (еще не опубликованы), дали представление о том, что помимо всего прочего модификация других сегментов может приводить к дестабилизации способности CENH3 связываться с ДНК.

С целью обнаружения мутированных генов CENH3 в различного вида растениях были созданы TILLING-популяции, характеризующиеся высоким уровнем мутаций, для кукурузы обыкновенной (*Zea mays*), рапса (*Brassica napus*), сорго двуцветного (*Sorghum bicolor*) и сахарной свеклы (*Beta vulgaris*) и проскринированы на содержание мутаций CENH3. Для этого, после создания ампликонов, охватывающих все экзоны генов CENH3, были проанализированы от 1000 до 10000 индивидуумов каждого вида растений посредством метода секвенирования по Сэнгеру. Кроме того, растения сахарной свеклы M2 тестировали на наличие мутаций с использованием специфичной ПЦР.

Кроме того, воздействие идентифицированной мутации внутри гена CENH3 на первичную и вторичную структуру закодированного белка оценивали, используя, помимо прочего, программное обеспечение Prof (Rost, B. и Sander, C. (1994a), “Объединение эволюционной информации и нейронных сетей для предсказания вторичной структуры белка”, *Proteins*, 19(1), 55-72; Rost, B. и Sander, C. (1994b), “Сохранение и предсказание доступности растворителю в белковых семействах”, *Proteins*, 20(3), 216-26; Rost, B., Cadafio, R., Fariselli, P. и Sander, C. (1995), “Трансмембранные спирали, предсказанные с 95% точностью”, *Protein Sci*, 4(3), 521-33). В Таблицах 9-12 показаны идентифицированные мутации в *B. napus*, *Z. mays*, *S. Bicolor* и *B. vulgaris*, которые разделены на мутации, вызывающие ошибку сплайсинга, и мутации, вызывающие аминокислотную замену. Мутация в

сайте сплайсинга представляет особый интерес. Такая мутация(и) может вызывать нарушение в сплайсинге (ошибка сплайсинга), что приводит к увеличению трансляционной клеточной продукции не полностью функционирующих белков CENH3, т.е. белков с нарушенной стабильностью, уменьшенной способностью связываться с ДНК, измененной геометрической формой, предпочтительно измененной вторичной или третичной структурой, либо с беспорядочным фолдингом, по сравнению полностью функционирующим белком CENH3 дикого типа. Растения с геномом, который был гетерозиготным в отношении такой мутации(й), были жизнеспособными.

**Таблица 9.** Мутация CENH3 из *Brassica napus* (аа: аминокислота; nd: не определена; у: да; н: нет). Замена аминокислоты показана как X#Y, т.е. аминокислота X (однобуквенный символ) заменена аминокислотой Y в положении #.

идентификатор мутации ( <i>Brassica napus</i> )	кодон дикого типа	мутантный кодон	мутация	случайность во вторичной структуре
BN_CenH3_01			ошибка сплайсинга после аа в положении 18	nd
BN_CenH3_02			ошибка сплайсинга после аа в положении 33	nd
BN_CenH3_03			ошибка сплайсинга после аа в положении 37	nd
BN_CenH3_04			ошибка сплайсинга после аа в положении 37	nd
BN_CenH3_05			ошибка сплайсинга после аа в положении 163	nd
BN_CenH3_06	tcc	ttc	S9F	у
BN_CenH3_07	cga	caa	R16Q	у
BN_CenH3_08	tcg	ttg	S24L	у
BN_CenH3_09	gaa	aaa	E29K	н

BN_CenH3_10	ggt	gat	G30D	n
BN_CenH3_11	gcg	acg	A33T	n
BN_CenH3_12	ccg	ctg	P35L	y
BN_CenH3_13	agc	aac	S41N	n
BN_CenH3_14	gga	gaa	G43E	y
BN_CenH3_15	cct	tct	P50S	n
BN_CenH3_16	cca	cta	P55L	n
BN_CenH3_17	ggt	gat	G57D	n
BN_CenH3_18	gga	gaa	G61E	y
BN_CenH3_19	cga	caa	R65Q	y
BN_CenH3_20	cga	tga	R65stop	n
BN_CenH3_21	cct	tct	P71S	y
BN_CenH3_22	gcc	acc	A105T	y
BN_CenH3_23	cga	caa	R110Q	y
BN_CenH3_25	agt	att	S114N	y
BN_CenH3_26	cct	tct	P121S	n
BN_CenH3_27	tgg	tga	W127stop	n
BN_CenH3_28	ctt	ttt	L132F	y
BN_CenH3_29	gcg	acg	A138T	n
BN_CenH3_30	tgc	tac	C153Y	y
BN_CenH3_31	gct	gtt	A154V	y
BN_CenH3_32	cgt	cat	R159H	n

BN_CenH3_33	gtt	att	V160I	n
BN_CenH3_34	gat	aat	D166N	n
BN_CenH3_35	gag	aag	E168K	n
BN_CenH3_36	cgt	cat	R172H	n
BN_CenH3_37	ctt	ttt	L173F	n
BN_CenH3_38	gga	gaa	G174E	y
BN_CenH3_39	aga	aaa	R178K	n

**Таблица 10.** Мутация CENH3 из *Zea mays* (aa: аминокислота; nd: не определена; y: да; n: нет). Замена аминокислоты показана как X#Y, т.е. аминокислота X (однобуквенный символ) заменена аминокислотой Y в положении #.

идентификатор мутации ( <i>Zea mays</i> )	кодон дикого типа	мутантный кодон	мутация	случайность во вторичной структуре
ZM_CenH3_01			ошибка сплайсинга после aa в положении 89	nd
ZM_CenH3_02			ошибка сплайсинга после aa в положении 115	nd
ZM_CenH3_03			ошибка сплайсинга после aa в положении 141	nd
ZM_CenH3_04	gcg	acg	A32T	nd
ZM_CenH3_05	gaa	aaa	E35K	nd
ZM_CenH3_06	cca	tca	P56S	nd
ZM_CenH3_07	gca	aca	A107T	nd
ZM_CenH3_08	caa	taa	Q114stop	nd

ZM_CenH3_09	gga	gaa	G152E	nd
ZM_CenH3_10	cgt	cat	R155H	nd
ZM_CenH3_11	gtg	atg	V89M	nd
ZM_CenH3_12	aca	ata	T139I	nd

**Таблица 11.** Мутация CENH3 из *Sorghum bicolor* (aa: аминокислота; nd: не определена; у: да; н: нет). Замена аминокислоты показана как X#Y, т.е. аминокислота X (однобуквенный символ) заменена аминокислотой Y в положении #.

идентификатор мутации ( <i>S. bicolor</i> )	кодон дикого типа	мутантный кодон	мутация	случайность во вторичной структуре
SB_CenH3_01			ошибка сплайсинга после aa в положении 26	nd
SB_CenH3_02	gca	gta	A62V	nd
SB_CenH3_03	act	agt	T64S	nd
SB_CenH3_04	gca	gta	A95V	nd
SB_CenH3_05	gca	aca	A25T	nd
SB_CenH3_06	tcg	ttg	S157L	nd

**Таблица 12.** Мутация CENH3 из *Beta vulgaris* (aa: аминокислота; nd: не определена; у: да; н: нет). Замена аминокислоты показана как X#Y, т.е. аминокислота X (однобуквенный символ) заменена аминокислотой Y в положении #.

идентификатор мутации ( <i>Beta vulgaris</i> )	кодон дикого типа	мутантный кодон	мутация	случайность во вторичной структуре
Bv_CENH3_01	gat	aat	D46N	nd

Bv_CENH3_02	gat	ggt	D46G	nd
Bv_CENH3_03	aga	aaa	A2K	nd
Bv_CENH3_04	ctg	cag	L106Q	nd
Bv_CENH3_05	ctt	cct	L109P	nd
Bv_CENH3_06	caa	cta	Q110L	nd

Кроме мутаций сайтов сплайсинга и точечных мутаций, вызывающих аминокислотные замены в аминокислотной последовательности белка CENH3, был идентифицирован кукурузный мутант (называемый Ми-мутант), содержащий инсерцию транспозона в 5'-нетранслируемой области гена CENH3 (см. SEQ ID NO. 24). Эта мутация вызывает расширение N-терминального хвостового домена. Таким образом, воздействие этой мутации очень схоже с воздействием мутации, описанной Ravi & Chan (2010), за исключением того, что эта мутация не является трансгенной.

### Тестирование мутантов CENH3

Для оценки биологической активности гаплоидного индуктора в идентифицированных мутантах и для тестирования гаплоиндуцирующей способности материнской и отцовской форм мутантные растения необходимо было скрещивать с другим растением-тестером того же вида. Предполагаемое гаплоидное потомство в результате такого скрещивания может быстро определяться, если используемые линии-тестеры несут рецессивную не-CENH3 мутацию. Значит, у гаплоидных растений проявляется рецессивный фенотип. Например, у кукурузы может проявляться мутация «глянцевость» (Мутанты кукурузы, Neuffer, MG и соавт. (1997), Cold Spring Harbor Laboratory, Нью-Йорк).

Результаты цитогенетического анализа участия индукторов в митозе и мейозе также свидетельствуют о пригодности мутантов в качестве гаплоидных индукторов. Гомозиготность определяли, используя молекулярные маркеры, полиморфизм тестера и потенциального индуктора. Гаплоидность как таковую тестировали цитогенетически.

При скрещиваниях с растениями-тестерами TILLING-растения с мутированным эндогенным CENH3 давали по меньшей мере 0,4% гаплоидного потомства. Часто, но не всегда, уровень индуцирования был выше, если при скрещивании тестер использовали в качестве материнского родителя.

Например, у *Brassica napus* в результате мутаций на основании аминокислотных замен в N-терминальном хвостовом домене уровни индуцирования составляют по меньшей мере 0,5% и частично достигают более 2%. Следовательно, местоположения мутаций не являются специфичными для определенной области этого домена, а скорее распределены по всему домену. Длина N-терминального хвостового домена у *Brassica napus* составляет 1-84 аминокислот. Мутации, придающие биологическую активность гаплоидного индуктора, могут, например, находиться в положениях 9, 16, 24, 29, 30, 33, 41, 43, 50, 55, 57 и 61, при этом не все эти мутации обязательно приводят к изменению вторичной структуры белка (рассчитано *in silico*). Сопоставимые результаты получены для более консервативного домена гистоновой складки, состоящего из трех спиралей и двух петель. Хотя во всем домене гистоновой складки могут обнаруживаться подходящие мутации, в среднем более высокие уровни индуцирования показывали  $\alpha$ 2-спираль, домен CATD и петля2. Исходя из этих данных изучения N-терминального хвостового домена и домена гистоновой складки, можно предположить, что и другие не тестированные положения и другие не тестированные аминокислотные замены будут придавать ту же или даже улучшенную гаплоидную индуктивность. Более того, другим видом модификации эндогенного гена CENH3 является замена нуклеотидов в сайтах сплайсинга, что в итоге приводит к ошибкам сплайсинга. Такие мутации также подходят для придания биологической активности гаплоидного индуктора. Наблюдаемые уровни индуцирования показывали по меньшей мере 0,5% гаплоидного потомства. Даже в этом случае можно предположить, что и другие не тестированные сайты сплайсинга будут придавать ту же или даже улучшенную гаплоидную индуктивность.

Например, у *Zea mays* в результате мутаций на основании аминокислотных замен в N-терминальном хвостовом домене уровни индуцирования составляют по меньшей мере 0,4%. Следовательно, местоположения мутаций не являются специфичными для определенной области этого домена, а скорее распределены по всему домену. Длина N-терминального хвостового домена у *Zea mays* составляет 1-62 аминокислот. Мутации, придающие биологическую активность гаплоидного индуктора,

могут, например, находиться в положениях 32, 35 и 56. Сопоставимые результаты получены для более консервативного домена гистоновой складки, состоящего из трех спиралей и двух петель. Исходя из этих данных изучения N-терминального хвостового домена и домена гистоновой складки, можно предположить, что и другие не тестированные положения и другие не тестированные аминокислотные замены будут придавать ту же или даже улучшенную гаплоидную индуктивность. Более того, другим видом модификации эндогенного гена CENH3 является замена нуклеотидов в сайтах сплайсинга, что в итоге приводит к ошибкам сплайсинга. Такие мутации также подходят для придания биологической активности гаплоидного индуктора. Наблюдаемые уровни индуцирования показывали по меньшей мере 0,4% гаплоидного потомства. Даже в этом случае можно предположить, что и другие не тестированные сайты сплайсинга будут придавать ту же или даже улучшенную гаплоидную индуктивность.

Ми-мутант, содержащий инсерцию транспозона в 5'-нетранслируемой области гена CENH3 (ID NO. 24) тестировали на биологическую активность гаплоидного индуктора. Эта нетрансгенная мутация порождает уровень индуцирования более 1%.

Кроме того, результаты скрещивания различных культур показывают, что идентифицированные и отмеченные мутации могут быть функциональными и в других видах растений. Таким образом, мутации могут вводиться в другие виды растений посредством таких технологий, как TILLING, мутагенез или геномное редактирование (например, нуклеазы CRISPR/Cas, TALEN, «цинковые пальцы» и т.д.). Более того, биологическую активность и эффективность гаплоидного индуктора может быть еще больше улучшена посредством объединения различных идентифицированных мутаций в одном растении и/или путем модификации генетического фона гаплоидного индуктора. Объединение различных мутаций может эффективно достигаться с помощью геномного редактирования, либо посредством создания вторичной мутации мутантного гаплоидного индуктора.

SEQUENCE LISTING

<110> KWS SAAT SE

<120> Haploid Inducer

<130> KWS0223PCT

<150> EP 14004389.4

<151> 2014-12-23

<160> 24

<170> PatentIn version 3.5

<210> 1

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> amino acid consensus sequence of the N-terminal tail domain of the CENH3 (part A)

<220>

<221> misc\_feature

<222> (7)..(8)

<223> Xaa can be any naturally occurring amino acid

<400> 1

Met Ala Arg Thr Lys His Xaa Xaa Ala Arg Arg Ser Arg Lys Arg  
1 5 10 15

<210> 2

<211> 14

<212> PRT

<213> Artificial Sequence

<220>

<223> amino acid consensus sequence of the N-terminal tail domain of the CENH3 (part B)

<220>

<221> misc\_feature

<222> (6)..(6)

<223> Xaa can be any naturally occurring amino acid

<400> 2

Gln Ser Gln Thr Gln Xaa Lys Lys Lys His Arg Tyr Arg Pro  
1 5 10

<210> 3

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> amino acid consensus sequence of the Alpha-N-helix of the CENH3

<220>  
<221> misc\_feature  
<222> (10)..(10)  
<223> Xaa can be any naturally occurring amino acid

<400> 3

Gly Thr Val Ala Leu Arg Glu Ile Arg Xaa Phe Gln Lys Thr Thr  
1 5 10 15

<210> 4  
<211> 11  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> amino acid consensus sequence of the Alpha-1-helix of the CENH3

<400> 4

Ala Ala Pro Phe Ile Arg Leu Val Arg Glu Ile  
1 5 10

<210> 5  
<211> 13  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> amino acid consensus sequence of the loop1 of the CENH3

<220>  
<221> misc\_feature  
<222> (7)..(7)  
<223> Xaa can be any naturally occurring amino acid

<400> 5

Thr Asn Phe Leu Ala Pro Xaa Glu Val Thr Arg Trp Thr  
1 5 10

<210> 6  
<211> 29  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> amino acid consensus sequence of the Alpha-2-helix of the CENH3

<400> 6

Ala Glu Ala Leu Leu Ala Leu Gln Glu Ala Ala Glu Asp Phe Leu Val  
1 5 10 15

His Leu Phe Glu Asp Ala Met Leu Cys Ala Ile His Ala  
20 25

<210> 7  
<211> 7  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> amino acid consensus sequence of the loop2 of the CENH3

<400> 7

Lys Arg Val Thr Leu Met Lys  
1 5

<210> 8  
<211> 10  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> amino acid consensus sequence of the Alpha-3-helix of the CENH3

<400> 8

Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly  
1 5 10

<210> 9  
<211> 6  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> amino acid consensus sequence of the C-terminal domain of the CENH3

<400> 9

Gly Lys Gly Arg Pro Trp  
1 5

<210> 10  
<211> 537  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Arabidopsis thaliana - cDNA coding for CENH3

<400> 10

atggcgagaa ccaagcatcg cgttaccagg tcacaacctc ggaatcaaac t	tgatgccgccc	60
gtgtgttcat cttctcaggc ggcaggtcca actacgaccc cgacaaggag agg	cgggtgaa	120
gttggagata atactcaaca aacaaatcct acaacttcac cagctactgg t	aaggaga	180
ggggctaaga gatccagaca ggctatgcca cgaggctcac agaagaagtgc tt	atcgatac	240
aggccaggaa ccgttgctct aaaagagatt cgccatttcc agaagcagac aa	acccttctt	300
attccggctg ccagttcat aagagaagtgc agaagtataa cccatatgtt gg	ccccctcccc	360

caaataatc gttggacagc tgaagctttt gttgcttttc aagaggcgcc agaagattac 420  
ttgggttgtt tggttctcaga ttcaatgctc tgtgctatcc atgcaagacg tgttactcta 480  
atgagaaaag actttgaact tgcacgccgg cttggaggaa aaggcagacc atggta 537

<210> 11  
<211> 178  
<212> PRT  
<213> Arabidopsis thaliana

<400> 11

Met Ala Arg Thr Lys His Arg Val Thr Arg Ser Gln Pro Arg Asn Gln  
1 5 10 15

Thr Asp Ala Ala Gly Ala Ser Ser Ser Gln Ala Ala Gly Pro Thr Thr  
20 25 30

Thr Pro Thr Arg Arg Gly Gly Glu Gly Asp Asn Thr Gln Gln Thr  
35 40 45

Asn Pro Thr Thr Ser Pro Ala Thr Gly Thr Arg Arg Gly Ala Lys Arg  
50 55 60

Ser Arg Gln Ala Met Pro Arg Gly Ser Gln Lys Lys Ser Tyr Arg Tyr  
65 70 75 80

Arg Pro Gly Thr Val Ala Leu Lys Glu Ile Arg His Phe Gln Lys Gln  
85 90 95

Thr Asn Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Glu Val Arg Ser  
100 105 110

Ile Thr His Met Leu Ala Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu  
115 120 125

Ala Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu  
130 135 140

Phe Ser Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu  
145 150 155 160

Met Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg  
165 170 175

Pro Trp

<210> 12

<211> 4001  
 <212> DNA  
 <213> Brassica napus

<220>  
 <221> Intron  
 <222> (1892)..(1972)

<220>  
 <221> Intron  
 <222> (2020)..(2104)

<220>  
 <221> Intron  
 <222> (2148)..(2208)

<220>  
 <221> Intron  
 <222> (2264)..(2358)

<220>  
 <221> Intron  
 <222> (2399)..(2507)

<220>  
 <221> Intron  
 <222> (2624)..(2755)

<220>  
 <221> Intron  
 <222> (2832)..(2933)

<220>  
 <221> Intron  
 <222> (3012)..(3091)

<400> 12

tgtccggag gatccaccgg cggtggtcgt tactccattt taacttgatg tcttcaaaggc	60
agaggacatg gtatggtggc ggcagaatct attttagtt gttaattttt ctttcctct	120
gattctttt attttttcg aatgaactaa ctgggggtt attcagaaga attatcatct	180
aaaaactgat tcaataaaca aaataattt catattcac aatgagccat tagtaaacaa	240
gtcgaaagtg aaaccaaatg ggaagagaac aatttaataaaaatgtt ctaatttcct	300
acttttatg aattgaactc ccgaagagaa tggccgaaga acggagtaaa agctcaatga	360
ttgtaaaagc tatgtattct ctgttttgc ggaaaaagct ttttgtttgc actcataggc	420
ctgatatgtt gtatggctc ttacatatt gggtcttttgc tggtcttatta aacggttact	480
gaagaattag ttatcgcat taaaaaaaaa atatttggaaa accacatatc taaaatctca	540
atatattatt tattgataga tgaaaagaaa aatagttat aaaataatat taattggtaa	600
atggaaagtca aaatttaca ttatatacat gactagttcg attctccggc tgaaacccat	660
ttaccactga gccaagacca cttgattata tatacaagc tcttgggggg tgcattaaaca	720
aaagttgcat cccaaatttc aaaacaacaa ctaaattaat gtctttctat tttcaagtt	780

ttatctccaa acctattaat agttaattt tttttttaa aattaggaaa ccggtaaaaa 840  
tatatttaa atacaaaaaa cttaaacaat gaatcattta ttatatttatt cttcaaatt 900  
taaatatccg aaccggccc aaaatatccg aacccgaaca taaaatatcc gaaccgact 960  
cgaagtgtag aaaatatccg aacgggtttt ataccttat actgaaatac cctatacgaa 1020  
cccgaatgtg tatccgaacg cccctacaa tatatgatca tcatttgtat cttgattgaa 1080  
aaaaaaaaaa gttaaactat tgatcacaaa atttcaatg tgagactttt accattttta 1140  
gtcattata gtcgtttta aaaattcaaa atataactta taagaaaaaa tctaattttt 1200  
tttattatat gcttaatgtg attgttaat ttctttaat aatataaaat taaacaaaaaa 1260  
atgagaggtt aaaaaatttg ttatcaaata tgtattattc ataatcatta attgtcatat 1320  
atatgttaat tatatttagt aatttcgtag ttttattta agaaaaagaaa aaaatattat 1380  
tttgacact actaattaat ttgatagtta gtttaataaa aaatatatta tattattata 1440  
tggaccaact tattttcta aaaaaacc actgtttaa aaaccaaacc aactataaac 1500  
cgagatata ccggatttag tggctaaaac actctttgtat tatatgtgct gagcaaacc 1560  
tctgagttag atggcgtgtt aagaagtagg aggaccattc atgccttta tgagttgtag 1620  
tctgtgtgt aaaaaagaa gcgttggtgt gaaagaaagc agaaggattt gaaaatcaaa 1680  
aaaattgaag gagaagcggg aaaacaaata atctctccct ccgcttttt ttctccaaat 1740  
aatcaatctc tcatttcatt ttttaaccca agttttgat aattatttca aagggttta 1800  
tttatctttt attcctccgg cggcagtaag tagtaatcaa tggcgagaac caaacatttc 1860  
gcttccaggg cacgagatcg caatcgaact agtttagtact ctctctctct ctgcctttt 1920  
tttgatattt attttctagg ttaaacccta atttggcatc tggaaattttt agatgcgact 1980  
gcttcatctt cggcggcggc ggccgaaggt ccgagtgcgg tacgtcatct attttctttt 2040  
cccgccccg gttttacgc aaatctcggtt actgttttt tgacgaatcg attgaaatgt 2100  
gtagaccccg acgagaagag aaggcagcca aggagaagct caacagagtg agtcttctt 2160  
tttcattttc tgagatccat gaatcctttt catctctcgt gtgttgtgac atgaatcaat 2220  
tgcagcagca actcctacta cgactccacc agccggtaga aaagtaagtt acattccat 2280  
ttcacaccat tcatttgctt ctttatcaac aaactgctct ctcattgtt tttttgttt 2340  
tgttttggtt ttgtgaagaa aggaggact aagcgaacta aacaagctat gcctaaaagt 2400  
tagtgacaga tttaaaatc tctattttgg atcatcattc tctcaggaca tgtctatttg 2460  
catttgttct tattatgtct gtctgtctgt cttgtcccc ctgttaggtt ccaacaagaa 2520  
gaagacattc cgttacaagc ctgaaaccgt tgccctcaga gagattcgcc atttccagaa 2580  
gaccaccaaa cttcttatcc ctgccgctag tttcatccga gaagtttagta atgaactttt 2640  
ttattcatac attcccgctt actgttttc aatgactctg caattactga tatagaattt 2700

ggagcaacca ttatgggtg atttctcaa ctacaatta ctaataactat cccaggttag	2760
aagtgtcacc cagatcttg cccctccga tgttaccgt tgactgctg aagctttat	2820
ggctattcaa gaggtacgtg tactccttcc ctctttgtt tcctatttc cacttgatgt	2880
ctaattaaa ctgatcgaaa ttttttata tttctttgg tgtggggcgg ggcaggcggc	2940
tgaagatttt ttaattggct tgttctctga tgctatgctt tgcgctatcc atgcaaggcg	3000
tgttactcta agtaagtagt actccccaaa ataagggaaac ccattttata tacaacattt	3060
cctcatccat gtctgcttct cttcatatca gtgagaaaaag atttttagct tgacgcccgt	3120
cttggaggaa aaggcagacc attgtgatcg tttcgaggt tgtataactt tgttcactcc	3180
ttatgtcttg tcatttgtga tctgactgac actttctttt gaaacataac tgcttgattt	3240
aatatctagg ctgtaaaact tatccctcct tgtttactat ctttatgtct ttcccttgg	3300
aattgatagt ttccatttgag atttcacttg cacgaaacat atctgctttc tcaatatctc	3360
tcagtcttag aaaggcgtat tgactaaaag aaaagaaaaat ttagaggaag atttgtaaag	3420
acatgtgttt agagaggcgt taattaaaaaa cacacgcttc tgctagcctt gctatttgat	3480
tcccaatttc aactttttc gaggcatatt ataaagttt taaatgtact tggcacttca	3540
actttataaa tttatataac gattttatttca taatagagca tttgtgattt catagtgttg	3600
tcatgaaact caagtaatttcc acaccgtccg atgttgctat tgctataataa aatgttgaaa	3660
aaattgtcaa aacagaacaa aaaacaacat agttgtctct atggtaaaa actatcacta	3720
agttgtctct atagtataat attttcgca atcccaaaac taattttct ttaatcaaataat	3780
taaacataaa ctAAAaccat tttaaaaag tttaatggaa aaagataaaaa aaataaggta	3840
atctcgtaat gttttaaaaaa ggaaaaaaaaa tgtaaaaaca attttaaaaaa aagaacacac	3900
gacacagatc aaaaatataca tgtaatctaa ttgcatttgg tttctaaaat cttccaaaac	3960
tattctttta aaattctcta aggtaaaact tgattccaaat a	4001

<210> 13  
 <211> 540  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Brassica napus - cDNA coding for CENH3

<400> 13	
atggcgagaa ccaaaccattt cgcttccagg gcacgagatc gcaatcgaaac taatgcgact	60
gcttcatctt cggcgccggc ggcggaaaggc ccgagtgcgaa ccccgacgag aagagaaggc	120
agccaaggag aagctcaaca gacaactcct actacgactc caccagccgg tagaaaaaaaaa	180
ggagggacta agcgaactaa acaagctatg cctaaaagtt ccaacaagaa gaagacattc	240
cgttacaagc ctggAACCGT tgccctcaga gagattcgcc atttccagaa gaccaccaaa	300

cttcttatcc ctgccgctag tttcatccga gaagtgagaa gtgtcaccca gatcttgcc 360  
cctcccgatg ttacccgttg gactgctgaa gctcttatgg ctattcaaga ggccggctgaa 420  
gatttttaa ttggcttggt ctctgatgct atgcttgcg ctatccatgc aaggcgtgtt 480  
actctaata gaaaagattt tgagcttgcg cgccgtctt gaggaaaagg cagaccattg 540

<210> 14  
<211> 180  
<212> PRT  
<213> Brassica napus

<400> 14

Met Ala Arg Thr Lys His Phe Ala Ser Arg Ala Arg Asp Arg Asn Arg  
1 5 10 15

Thr Asn Ala Thr Ala Ser Ser Ala Ala Ala Ala Glu Gly Pro Ser  
20 25 30

Ala Thr Pro Thr Arg Arg Glu Gly Ser Gln Gly Glu Ala Gln Gln Thr  
35 40 45

Thr Pro Thr Thr Pro Pro Ala Gly Arg Lys Lys Gly Gly Thr Lys  
50 55 60

Arg Thr Lys Gln Ala Met Pro Lys Ser Ser Asn Lys Lys Lys Thr Phe  
65 70 75 80

Arg Tyr Lys Pro Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln  
85 90 95

Lys Thr Thr Lys Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Glu Val  
100 105 110

Arg Ser Val Thr Gln Ile Phe Ala Pro Pro Asp Val Thr Arg Trp Thr  
115 120 125

Ala Glu Ala Leu Met Ala Ile Gln Glu Ala Ala Glu Asp Phe Leu Ile  
130 135 140

Gly Leu Phe Ser Asp Ala Met Leu Cys Ala Ile His Ala Arg Arg Val  
145 150 155 160

Thr Leu Met Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys  
165 170 175

Gly Arg Pro Leu  
180

<210> 15  
 <211> 3088  
 <212> DNA  
 <213> Sorghum bicolor

<220>  
 <221> Intron  
 <222> (377) .. (487)

<220>  
 <221> Intron  
 <222> (514) .. (618)

<220>  
 <221> Intron  
 <222> (662) .. (1094)

<220>  
 <221> Intron  
 <222> (1217) .. (1312)

<220>  
 <221> Intron  
 <222> (1391) .. (2087)

<220>  
 <221> Intron  
 <222> (2164) .. (2556)

<400> 15

catctctcac	tgccatccgg	gtccactact	cccaacgttc	ggcacgcccag	gtatagccgt	60
taccccgta	ggccccactg	gtacacggac	aaaggtagc	ggtcaccgcg	aatcgtgaat	120
acttgtgact	acgggggtgct	aattataaaa	acgcccaca	tccttcgtt	tcgccatttc	180
accccccattc	ccttcccgta	gagagaaaaa	aaacccaccg	tcgacccgccc	cggccgcccc	240
agagttctga	atcgaaaccg	tcggccgcga	ccgcgagagc	agcgcggggc	gcccaccgtg	300
atggctcgaa	ccaagcacca	ggccgtgagg	aagctgccgc	agaagccaa	gaagaagctc	360
cagttcgagc	gcgcaggtaa	gcccgctcc	ccgctgaa	ccccctccg	cctcgcgagc	420
agacgctgcc	gctgctctcc	gtcgccccctg	gtgctaagcg	cgttcctttt	tttccttct	480
tttgcaggtg	gggcgagtagc	gtcgccgacc	ccggtagtg	cgtgcgtgcg	tgcgaaatt	540
ggttttagcc	ctcctttgc	ggttcgcct	tttggggc	tggtctca	tgcttgcaat	600
ctgtttgatg	aatgcagga	gaggaggaat	gctgggaccg	ggggaggagc	cgcggctcgc	660
ggtgaggatc	tcttgcgt	tgctgggttt	ggaaatttcc	ggcgcgaaat	tatgtggatt	720
tctaggttta	tctgccgtct	ttcttcttgc	cttctctttt	ggctctgggg	tgagaagtta	780
gggtgggtgg	gcggacatgg	tgcgttattt	cgccgtatcg	tttgggttgg	tgcttctca	840
tccttttaat	tccaacatgc	cttgtaaaaaa	ttgcacaaga	tttggttttt	catgcatgtc	900
tcagtgttgc	taatttgcct	ttccggttcg	gttggtagaa	ttcaatttct	tggcgcaata	960
tgcatcttct	tttggttgcaa	catgagggcgc	aatgtgccag	ttccatatgg	gcgtcgccgt	1020

tttgaagtta ctaccttgct tgctcttcgt attataggcg tcattcacaa tagtatgttt 1080  
tcttggagat gcagttgcac gggggcgtgt ggagaagaag catcgctggc gggcagggac 1140  
tgtagcgctg cgggagatca ggaagtacca gaagtccact gagccgctca tccccttgc 1200  
gcccttcgta cgtgtggtgg gtgcacatctt taccaattgt tgtccactcc atagaatggg 1260  
tttggtctgc agtctgtctg atggaaagtt attcttctga gaaaaaatgc aggtcaaaga 1320  
gttaactgca ttcataaacag actggaggat agggcgctac acccctgaag ccctccttgc 1380  
gctgcaagag gtcagttatg aaacatgtct tgtgtatcag ttaagatcat cttctataga 1440  
cataattgtt atcatgaagt cttttctgt taatcggtct ggtactactt aataatcagg 1500  
atttcagatt gctgccttc ctatgtgtgt agtcaaaagg gaatttaagt gctgttaggt 1560  
actgtttgtt ttgggtgttt gaaccctgcc gcgatcggtt gttgttattc catgtttgtt 1620  
tctgtggcag cggacgttca cggtagatg ggatacgggc gtgtgaaaca tagttacgg 1680  
ccatcttcat ggcttatcca ttacgctgc tcgtccgctc acttgttattg tgccggcaacc 1740  
aaactttgt tactagtgtt actggtagcg ttgcaaatct ttccatttgc gttaccactc 1800  
cctatgggag ccaaacagca ccttagtgtt gattccattt gtattactt agctagcttc 1860  
cttgctattt gtcgcctcgat tgtactgtt tgatcgaagt gctgaaaact ttgtgcctg 1920  
catagcatga ttagagaact ttagttaca ttattcaat accttaagac tgcatcgat 1980  
atagataaat tattttcct aattgttctg gttaactgtt ttaggtttcc atattttgt 2040  
atgtgtatca tttaaattat tttttttttt ttcctccctg tctacaggca gcagaattcc 2100  
acttgataga actgtttgaa gtggcgaatc tgtgtgccat ccatgccaag cgctgttac 2160  
tcagtaagtt atcactgaat gaactccttt tcctctgtac tattacgcct aatggagatg 2220  
tgtgtatgcat ttgggttac acgattctt agtgattctg cttcagttgg atatgataaa 2280  
tcttagatgtt atttaaagtg gcaaattgt tacgagtgg aatagtaatg ttcaaatagt 2340  
gaaaagtgcattttt taataggcca ttatatggtt tgattgtcaa caaatgcattttt 2400  
aagaaatagt aaatattata acagttatgg ctttagagagt ggacaaaaaa tcggtaatgg 2460  
tgagctttgtt ataaacacta aaactggctg agaaaatctga taactcaagg atctatagga 2520  
aatgtattat cctaaatgtt ttccctccctg ctgcagtgc aaaggacata caacttgcaa 2580  
ggcgtatcgg aggaaggcgt tggcgtatgc atccattctg attctgatta cttgttcgg 2640  
gtggaatttg tttagaggag ttagacatta gtcttggta atgctgtgc tggttcctaa 2700  
tctgtttcac agtttagtggg ctcttctggg atgatctgtt aacacctgtg gagttatgtt 2760  
tgttaggaaac acctgaactg aacaacccaa agttgtttt gttgctcttc aaccatttgtt 2820  
ttgcttcaga gatcgattct aaactgcattg ctaattagtc tatggttgaa caaaaatttat 2880  
caaataaaaa tgaaagtgtt atagtagcaa aatccaaaaaa aaaaaggatc caaacaaggc 2940

ctaaaatcat ggttcttct cctttgaac tgggtgcaag tatggacagg cacagaagaa 3000  
aacgcctag caaaccgtt gttttttt cttcgttgta ccacacgaca ctgttcgttc 3060  
ctagttgcgc ctttttgtg tagaagt 3088

<210> 16  
<211> 471  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Sorghum bicolor - cDNA coding for CENH3

<400> 16  
atggctcgaa ccaagcacca ggcgtgagg aagctgccgc agaagccaa gaagaagctc 60  
cagttcgagc gcgcagggtgg ggcgagtacg tcggcgaccc cggagaggag gaatgctgg 120  
accgggggag gagccgcggc tcgcgttgca cggggcggtg tggagaagaa gcatcgctgg 180  
cggcaggga ctgtagcgct gcgggagatc aggaagtacc agaagtccac tgagccgctc 240  
atcccctttgcgcccatttgcgt acgtgtggtc aaagagttaa ctgcattcat aacagactgg 300  
aggatagggc gctacacccc tgaagccctc cttgcgtgc aagaggcagc agaattccac 360  
ttgatagaac tgtttgaagt ggcgaatctg tgtgccatcc atgccaagcg cgtaacagtc 420  
atgcaaaagg acataacaact tgcaaggcgt atcggaggaa ggcgttggtc g 471

<210> 17  
<211> 157  
<212> PRT  
<213> Sorghum bicolor

<400> 17

Met Ala Arg Thr Lys His Gln Ala Val Arg Lys Leu Pro Gln Lys Pro  
1 5 10 15

Lys Lys Lys Leu Gln Phe Glu Arg Ala Gly Gly Ala Ser Thr Ser Ala  
20 25 30

Thr Pro Glu Arg Arg Asn Ala Gly Thr Gly Gly Ala Ala Ala Arg  
35 40 45

Val Ala Arg Gly Arg Val Glu Lys Lys His Arg Trp Arg Ala Gly Thr  
50 55 60

Val Ala Leu Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr Glu Pro Leu  
65 70 75 80

Ile Pro Phe Ala Pro Phe Val Arg Val Val Lys Glu Leu Thr Ala Phe  
85 90 95

Ile Thr Asp Trp Arg Ile Gly Arg Tyr Thr Pro Glu Ala Leu Leu Ala  
100 105 110

Leu Gln Glu Ala Ala Glu Phe His Leu Ile Glu Leu Phe Glu Val Ala  
115 120 125

Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Val Met Gln Lys Asp  
130 135 140

Ile Gln Leu Ala Arg Arg Ile Gly Gly Arg Arg Trp Ser  
145 150 155

<210> 18  
<211> 5834  
<212> DNA  
<213> Zea mays

<220>  
<221> Intron  
<222> (1820 )..(1917)

<220>  
<221> Intron  
<222> (1944 )..(2035)

<220>  
<221> Intron  
<222> (2085 )..(2239)

<220>  
<221> Intron  
<222> (2356 )..(2446)

<220>  
<221> Intron  
<222> (2525 )..(4530)

<220>  
<221> Intron  
<222> (4607 )..(5346)

<400> 18  
ttatgttagag gcaattgcag tagtgccctg gtttttagagt gtaactacag atttgtccct 60  
attttttag agtttgcgtg tttgtccctg ttttttcaaa tcaaactatt gtataccct  
actccattag ttatacttaa caatgttaag tcttgataaa aagacaaggg ataattggat 120  
tagtgaccct gtttttagagt gtaattatag ctttgcccga tgttttagac ttcacatgtt  
tttatgacaa ttcaaattgt ttccataaca tcttaaatta ttttgacaac atttagaatt 180  
gttttgcaat aatttaaatt attccaaaaaaa taaaaaatatt ttgacaatta ttttatcaac  
aaattaaatt attttttta caaaaataatt tgtcaaggta ctttttaaa attttgaaaa 240  
taatcaaatt attgtaaata taatttgaat tggtcgaaa ataatttgaa ttgtcataaa 300  
360  
420  
480

aacacgtaag tctaaaaatt aggcgtaaaa ctacaattat actctaaaaa ggtggtaatg 540  
gcgttagttgt tccttgcata ttttatcaag acatagcacc gtgcagtaca actaatggag 600  
tagtgacata caacaatttt gtttaaata atagggtaa atacgcaaag tctaaaaaac 660  
agaggtaaat ctacaattac atgttaaaat agaggcatcg atacaattgt acctttata 720  
tagcagcatg cgcctgttg ggatacaatt gtaccttca catgtcttct agatggttcc 780  
caacccttg gccaaagatcg tacagataat attgcgagga gcccaaatca acgggtgtcca 840  
tatgttatgt tgatgtggat ggttaccta ggcgaaaag tgcgctgggt tcgtccgtac 900  
aaatatactt taagtatggt tttgattttt ttctatttt catttttaa ataaaacgag 960  
acaatcaaat ctgatataaa aatcaaatga attataaata gagacggaaa gagtatata 1020  
atttgttttgcattattta aagtattaaa agatagtgga cgaatgaacg tcctctatgt 1080  
ttaaaaagaac gtttagagg acgttgtgtt gttgaaggaa atatgaaaaa aaaatcttct 1140  
gcataatggtag aagggaggag cgtttacaca ttacttcgg gacttcaacc caaatatgtc 1200  
aagggttgtagtggctcag tgccggaaaaaa aaatccata tataccagat gtaaacacta 1260  
tcctttacag cctatcacat tcacatggtag aggttccaa agatagatca aaatttataa 1320  
aataatcatt taatattttt tttatttat ttatatggat aagcagctgg tgtatgtgag 1380  
gagctgtaaa agatattttt tacatccgag atgtaaagat ttttttaac tcaatgctgg 1440  
ttaccggctg ggaggacat gataaagaaa gcatctca ctgcattccg ggcccactac 1500  
tcaaacgttc ggcacgcccag gttggcaggt agccgttaca tcgataggca ctcggccact 1560  
cgcacgcaga caccacacca gtgtgctcag tgctcactgc tcaccataat aacgctgcac 1620  
ctctttcat ttcaccatct cctgccccct taaaaaaaaaag actcaccgtc gacacgcccct 1680  
cccgtcccgaa gagttctgaa tcgaaaccgt cggccacgag agcagtgcga ggcccacc 1740  
gcgatggctc gaaccaagca ccaggccgtg aggaagacgg cggagaagcc caagaagaag 1800  
ctccagttcg agcgctcagg taacccgggt cccgcgtcc ccccgcttc gcaaggagac 1860  
gctgtcgctt ctctccgacc ctggcgtctaa gcacgttcct tggccgtct tttgcaggtg 1920  
gtgcgagttac ctcggcgacg ccggtgagcg cgtgcgtgcg gggatcagtt ccctcctttt 1980  
gcctttttt gttggcgtgc tcttacttgc ttgcaagctg tttgatggaa tgcaggaaag 2040  
ggctgctggg accggggaa gagcggcgac tggaggtgac tcaggtgagg acctattgt 2100  
cgttgctgga tgctgggttt cgcttgcaat ctaattttgt tgcaagatga gggcgaatgt 2160  
gccagttcca tgtgggtgac atggtctcg agttactacc ttaattgctc accatagtt 2220  
gttttcttaa aaaaaacagt taagaagacg aaaccacgac accgctggcg gccaggact 2280  
gtagcgctgc gggagatcag gaagtaccag aagtccactg aaccgctcat cccctttgcg 2340  
ccttcgtcc gtgtgggggg tgcaggcgtg tttgtcctct gcatagtagtgggttcc 2400

gcattctgtc taatggaaag ttattcttct gagaaaaaaaa atgcaggtga gggagttAAC 2460  
caatttcgta acaaacggga aagtagagcg ctataccgca gaagccCTCC ttgcgctgca 2520  
agaggcagt tatgaaaaat gtcttatctc tctgttaaga tcctcttcat atacatagtt 2580  
gctattgcta tcgtgaagtc tttttttct gttaattggc ctggtaactac ttactagtca 2640  
ggatttcata ttgcggTTT tcctagtggc gtgttagttAA aaagtagttt aattgcttt 2700  
agttaaaagg ggtgttcagg gctaaagatc aactatgaga aaacagaaaAT tttcccaATT 2760  
cgataccCGA cagcattatG gcctgcgcta atggaggtgt ttccgggCAA atactctAGC 2820  
ctacctggGA agtaccttGG gttccccCTT catttcagGA aagtaaaaAG gaatgatCTT 2880  
caacctctAA tcgaaaaat caacaacagg ctggcCTtGC tgaaaggCA agatgttGTC 2940  
caaggctggT atagaaaACTC ttgtaaaATC gatgctatCC gcacaaccaa tctaccatCT 3000  
aatggTTTT ccacctcaAA aatggctgCT gcaaacaATT gacAAaaATAC gaagaaACTT 3060  
cctgtggaga gggagcaATC cagaagtttG cagcggggGT cactgcctCG tcaactggCC 3120  
cgtaacttGC ctcccaAAAGA acaaggGAgg tcttggAAatt ctggaccttG atcgTTTGC 3180  
gagggggcTA agactaAGAT ggctgtggCT acgatgGAAG agcaaAGATA gggcgtggAC 3240  
tgccttGAAG cttccttGtG acaAAactGA tgaagatCTC ttcaatgCTT ccacaactGT 3300  
cacggtagGC aatggAAAGA tagctGAatt ctggAAattCT agttggatCC aaggccaAGC 3360  
ccctaAGAAAC attgcGCCAA cactgttCAA gaaggAAAAG aggaAGAAACA tcacggtcGC 3420  
caaAGCGCTC actaacaACA attggattCG tttatgCTA ccataCACGG gtgagggggGA 3480  
gtttagagAG gtcgtctCTC tttggcaggC cataggtAAC atgcaAGAGC ttaacggTTT 3540  
ggaagacaAC atctcttGGA gatggacGGC agatggcAG tacagtGCTA gcagtGcATA 3600  
caaaatccAG ttCGcatCCA atttcaCTAA aatgaacCTC tgccCTtATT ggaaggCTAA 3660  
agtggAACCG aaatGCCGAT ttttgcttG gacactACTT cataAGAGAA ttctgactGC 3720  
cgataacCTT cataAAAGAG gttGcaACTC agcctcAGAA acaattcccc acttatGCAA 3780  
ggattgcccc ttttagtagAG aggtgtggAA caaagtttG tctcgggCCA actttccTTT 3840  
actgactGGG tctcccAGtG acacttCTT gtatgattGG tggacggACA tgtgcAGCCT 3900  
ttgcAGcAGA caggcaAGAA gaggttCGA cggctcgta tttcaCTTT ggtggAACTT 3960  
atggctggAA agaaATAACA gaatCTTCA aaggcAGcGT agaagtGtAG atcaagttGC 4020  
tctggcAGtC aaggattATG cttagtagCTG aagtctAGtT ggtttggACT agtggTTTG 4080  
ttgcTTTCT tttaatttC ttttagtTC tttttagtTT gtttgcTTT ccttaagttG 4140  
cttggagtCT gtattatCCT cttcttCTA atatAGatCG gagcGacAAA CCTTTGCC 4200  
cttccttCA aaaaaaAGTT aaaaggGAAT ttaactGCTT tcctagtggT gtagttaAAA 4260  
tggatttCAT attgcggcCT ttccTAGCTT gcttgcttatt gattggactA tagtGatCCA 4320

aatgctgata	actttgtcgc	tttgttaggc	atggtagag	agcttagagt	ttgcatttat	4380
tcaataccctt	gagactgcat	ttcatataca	taaattatttc	atgattattt	cttttctcta	4440
tttggctgg	ttaattaaga	gttttaggtt	tccatatttt	tgtacgtgca	tcatttaaat	4500
tcttgatttgc	ttttcggttc	ttgtctacag	gcagcagaat	tccacttgat	agaactgttt	4560
gaaatggcga	atctgtgtgc	catccatgcc	aagcgtgtca	caatcagtaa	gttattcactg	4620
agtgaactcc	ttttctctg	tagcattact	cctaataatgaa	atgtgtgatg	cattttgggt	4680
gcacgattct	ttagtgattc	tgcttcagat	ggatatgata	aatctagatg	ttatttgaa	4740
gtggcgaatt	gcttacgagc	ggaaatagta	atgttcaa	atgcgcaa	gcaactgttg	4800
acttttagta	ggccatttat	atggtttgc	taccaacaaa	tacgtcaatc	atatgatttgc	4860
attatcaaca	aaggaatcag	ctatatggtt	tgattatcaa	caaaggaaatc	agcttaggttt	4920
gcttatcaac	attcaacaaa	ggcatcaagt	aataactccat	ccgtttcaat	ttataattcg	4980
tttgactttt	tttatctaag	tttgatcg	tcgacttatt	aaaaaaaatc	ataattatttgc	5040
ttaattttgc	ttgtgatatt	gttttagtata	atatactta	aatgtgactt	tgagttttc	5100
atttttcg	aaaaaaaaat	gaataggacg	agccggtaa	acgtgacaca	aaaaagtcaa	5160
acgaattata	atttgggaca	cacggagtag	taaataatgt	aacaacttag	agagtgggac	5220
aaaaaaaaatct	ctagtggtgc	taaatttagt	tcagcttgc	ataaacacaa	gcattgatttgc	5280
agaaaatctga	caactcaagg	atctgttaga	aatgtttac	cctaaatgtt	ttccttactg	5340
atgcagtgc	aaaggacata	caacttgcaa	ggcgtatcgg	aggaaggcgt	tggcatgt	5400
atataatatc	cattctgatt	gcatcattct	tgtgaatttgc	ttttaggag	ctagacatttgc	5460
gtgttgttgc	atgctgcatg	gttccta	ctttcgcag	tctaacatct	gtggagtttgc	5520
tatgttacat	ggcaacagct	gaacatctgt	ggactatatg	gcaacagccg	aagattgtgt	5580
ctgtggata	actggttgtt	ttgggtgctc	ttagtagtt	tgttgcttc	aggtaaccat	5640
gctgcgaact	atgatgtttt	cattctcggt	ttgcttcagc	taaccgagat	cgattcagtc	5700
tgcagtatgg	actatggagt	aaactgc	ctgaaacccg	aaccactgct	gaaactgc	5760
gctgaaaccc	gaaccactgc	tacggcagtt	gccaggatag	caggagggcc	tttatgcaca	5820
gtgaaattga	gttag					5834

<210> 19  
<211> 471  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Zea mays - cDNA coding for CENH3

<400> 19  
atggctcgaa ccaagcacca ggccgtgagg aagacggcgg agaagccaa gaagaagctc

cagttcgagc gctcagggtgg tgcgagtacc tcggcgacgc cgaaagggc tgctggacc 120  
gggggaagag cggcgctgg aggtgactca gttaagaaga cgaaaccacg ccaccgctgg 180  
cggccaggga ctgtagcgt gcgggagatc aggaagtacc agaagtccac tgaaccgctc 240  
atccccttg cgccttcgt ccgtgtggtg agggagttaa ccaatttcgt aacaaacggg 300  
aaagtagagc gctataccgc agaagccctc cttgcgctgc aagaggcagc agaattccac 360  
ttgatagaac tgttgaaat ggcgaatctg tgtgccatcc atgccaagcg tgtcacaatc 420  
atgcaaaaagg acataacaact tgcaaggcgt atcggagggaa ggcgttggc a 471

<210> 20  
<211> 157  
<212> PRT  
<213> Zea mays

<400> 20

Met Ala Arg Thr Lys His Gln Ala Val Arg Lys Thr Ala Glu Lys Pro  
1 5 10 15

Lys Lys Lys Leu Gln Phe Glu Arg Ser Gly Gly Ala Ser Thr Ser Ala  
20 25 30

Thr Pro Glu Arg Ala Ala Gly Thr Gly Gly Arg Ala Ala Ser Gly Gly  
35 40 45

Asp Ser Val Lys Lys Thr Lys Pro Arg His Arg Trp Arg Pro Gly Thr  
50 55 60

Val Ala Leu Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr Glu Pro Leu  
65 70 75 80

Ile Pro Phe Ala Pro Phe Val Arg Val Val Arg Glu Leu Thr Asn Phe  
85 90 95

Val Thr Asn Gly Lys Val Glu Arg Tyr Thr Ala Glu Ala Leu Leu Ala  
100 105 110

Leu Gln Glu Ala Ala Glu Phe His Leu Ile Glu Leu Phe Glu Met Ala  
115 120 125

Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile Met Gln Lys Asp  
130 135 140

Ile Gln Leu Ala Arg Arg Ile Gly Gly Arg Arg Trp Ala  
145 150 155

<210> 21  
<211> 8441

<212> DNA  
 <213> Beta vulgaris

<220>  
 <221> Intron  
 <222> (99 )..(203)

<220>  
 <221> Intron  
 <222> (245 )..(331)

<220>  
 <221> Intron  
 <222> (344) ..(488)

<220>  
 <221> Intron  
 <222> (591) ..(3263)

<220>  
 <221> Intron  
 <222> (3339) ..(4377)

<220>  
 <221> Intron  
 <222> (4458) ..(8130)

<400> 21

ctactctttc tctctcttc	tctctccatt tctgttgaa atcatgagag	ttaaacacac	60
tgctgccagg aaatcaacca	ccaacggtcc tcgttcaagt tagttcctc	tctcttcttc	120
ttttttgttc gcattctctc	aatctatatt tcaaatttga aaaaaattgt	gatgctcata	180
aaccctaaaa ttttcttgc	cagaggctca gaaatctccg	cgcagtttgc aatcaccaca	240
atcggttctc tctttgtact	tttgatttgt ttcccttca	tttgcgtat gaatggctct	300
taattgtctt ttatTTactt	gaaaatttgc	gccttcttagt agttcaaaggc	360
cagaaacact gatgcaactc	ctcaaagtaa ctttttcttt	aatatttaggt ttaattttac	420
tgctgttgc caaattctgt	tggaaatttgc	aaatattttt ttctttaat ttgacggttt	480
cagagaagaa ggcttaccgc	cgtaagccgg	gcactgtggc actctggaa atacgcaa	540
ttcagaagtc attcaagccc	ttgattcctg	ctgcgcctt cattcaaca gtatgtattt	600
tttttgcgtt tacttaataa	atgaattttt	gactgggtt tgtgtggctg catagaata	660
tatttccata caactgaaat	tgccttagga	ggtatcgatg aatgtttgc acaaaataaa	720
taaatataag tgattatatac	ttgttaaaaaa	gccattataa ttgcaactta tatgtatgtt	780
gtaatgaggt caactagcta	tttggc	agtcacccac actttaacat aatTTTGTG	840
tctcgtaacc ttaaaaaaaat	ataagtaaag	ggttggatttgcgtat tagtgc	900
acccaatttag attgaaacat	aaggtaa	cagggtggta tcagcttcaa ttagatctaa	960
taagtgcagt ttatTTAGC	ttcggtgaaa	tgaacacacc cttaaagata gaaaatcgac	1020
actatataatg gtcctttta	gatatgatag	ttcgatattc tttttgggg tgtgttgaat	1080

gattaaatgg agtggtaat agctgatggg aactagagaa gatgctcagt agacagttat 1140  
tgtggagact atattactga ttacccctgt ttctgagtgg ttaggacaat gtgacaattg 1200  
atttggta ttatggtag atgttttct tttgttaaa agtgccaaga taggtgtgca 1260  
gttgctgatt ctcagttgc taagaattag ctgtgtctgt atttcgtacc tcagttgatt 1320  
ctaagtgaac atttcttga attgatgctt tgttcttgca tcatgcaact tggtaagct 1380  
ttctttagt tgctccagtg gcaatctagt ctggtatgtt tagaactctt gtgatggat 1440  
gagttcatca agatattgg gatccaatta gcctaaccaa tgtttttac cccctattgt 1500  
cactgactta tactccctta tctataaaaa taattgtgac attgatccat ctcctcacaa 1560  
tcattaatat tatatgtact gaccatctt acactctaa cactgaatct aagtagggga 1620  
atttggaa attcaatgt gaactagtagc acccttcttc ccaataatat tggtgacatt 1680  
tttattttga tttgtccat attgtcctct ttggtaattt aatgtatatt caccaattt 1740  
tctttcaat acccactata ccaacatata attggtaat tcattttat taattatatt 1800  
tcctaagagc ttgttgtgta aacgtggatg aattttagg catggatgaa gtattgttat 1860  
aatgaggta caacattact taatttcgaa ctgaggaca gagggatatg atgataaaac 1920  
aactttgct tgcttcttaa actcagaaga tagggttac accaagtggc atgtaaaagt 1980  
cactagatga ttatctatta caaggcctg tacaatctga agtacgatag gattgaagt 2040  
taacaacatt catcgaaagc tcataacttg tccttatatc aatataagtt gctggcatgt 2100  
gaaattgcgt tgcaagcatc catgagctag ctcaactatt aactattaaa ctttatattt 2160  
ttgcttgatc tagtatgagt cctactattt agttctcca tctaccttaa tatgtcgcat 2220  
acaccaacta atcattatcg ctagaatcaa taaacaaagc tttcttcct taggtgtatt 2280  
agtacctagc tcctgtataa ccaagagcac ccaaattggg aagaaaaagt agaattggct 2340  
catatctcta atcctacatt gatcattgaa aaggaccta agttctcat actgaaacat 2400  
catcttttg agcaggatat ctacgtagac gacaagaaag actactttgg ttgcccgtgc 2460  
attingtgc atcagacaac ttcttacca ctgtctaacf gcttgctttg gccatattgt 2520  
ggtcttctat gccaaaatta atgatatttc ttggcaccgc gctaattgata ttactgaatg 2580  
cgatgatcgt acgatttagaa ttattcaaa gtaggttagca attacttagtt ttgagcattg 2640  
agttcaata attagtaat taagtgtcaa acttgtacat ttggctaca tgtatggaa 2700  
ttagaattgg tacgaggaaa tatagcaaca ttacggcaa tggtcactca agtagaaagcc 2760  
attacatcaa atagtactag ttgaagtatt agttctcata atactaatca ttgtcattaa 2820  
tggaaatattg gaacgtaaat gccttaagg tgctgttagt ttagtagaaa ttctactatt 2880  
ctagtagat aatgcaattt attgaaactg ttgttaagat agcttggatc ccacatcgt 2940  
cttgcgtcta aataaatgga tgccataat cttctaattt ttaatttgat tctcttacca 3000

aacgagaaaa aataggagaa atccaatttg catgacctca ataaggaaat gttgttaatg 3060  
tgtgatgctt gtttctcatt tatagtctag agagagttat catgtccaag attgcagtct 3120  
tggtaactgag aaagttttagt tgggggttgg ctgcttcttgc agcctctttt ttttagagta 3180  
agacacttcc tagatataat tttctttatt ttttgtaaa ttccatatat actactacat 3240  
taacaggttt aagtttatat taggtgagag agattactca ccagtttgct ccttatgttg 3300  
gtcgttggca agctgaagct ctgatggccc ttcaagaggt gcagaccaac tcttttagcc 3360  
tttttttttc tggcatgtca agtgtggcta ttagattttc tgtgtgattc tcactccat 3420  
atatctatat atgtacatat taaagcacat tgataccat cttgtcagat gtggctttt 3480  
caattctttt ctaagttgag attttctct tggcgtaga tatgctcctg ccgaaatata 3540  
ctgctgtctt gttatccatc atgacttggt tatgcttgta tctggcatt atcttggcat 3600  
gcttaaaaac aagtattgaa cgagcctcct attgataaat ttactatta atattggatg 3660  
gcttctcaaa ttctaattggc agtgagatac tgttaagttg ggagaaatag attaagaaac 3720  
agaaaagatgt ttaccatgag agcaattgaa atagaaaata gagtaacttt ttgcaaagat 3780  
tttggtcctt tagatttttg aatactacct gataatgaag cattttctaa atttatgtgc 3840  
tttctatcta tcagatactg gaatacaatc aaattcctat cacgtactga gcattgtat 3900  
cagattcttgc tttgcttcct atcacatact ggaatcctt tgattgttga atacaagat 3960  
aatgaagcat tctctaactt tatgtgctgt aactactat aatgattctt gcttgccctcc 4020  
tatcacgtac tcaaattcctt tgtttgattt gtctcttata agaggaactt cctgtctttc 4080  
tttgcatttgc ctttagtattt atagaggtgc caacttatgg ctttgacaac tgaagctttt 4140  
atgcaaactc cggattttgt tggatggaaatg acaagtaaca cttagcatg tggattcagg 4200  
tcttaacggtt aagactttt aatgaatgtt ttaactgttag tagtttattt atataaaaaaa 4260  
agtggctctt caaactttt atgagatcat atcgaagtaa tcaaattttt gattcaggtt 4320  
cttctgctat tattttttgtt taagcatgtt ctatttttttga cagtctgtca attgtaggct 4380  
gcagagaatt ttattgtccg tttgtttgaa gatggatgc tttgtgcaat tcatgccaaa 4440  
cgagttacac tcagtggat tctgatttcc ttccgggttgc ctgttattat gcattatata 4500  
cactttgcctt caatatcgat atataaggag tccttggggcatattttt tggatgtttt 4560  
gttatatcctt gtttagtgcc tgctgcagtt gtgaacttac ggcctgtttt attagtggtt 4620  
ataaaatgtt gtaataactaa tataattttt tataaattttt taaaaaaaaat gctaataatca 4680  
atattttatggttaatgaaattttt ttagtataaa acatgagttc tcttttataaa gttttcattttt 4740  
ctatccaata ccacccccc aagtggtaat gaacggtaat gaaattttttagt gaagaaaatg 4800  
gatattttggg gatttagatg cattaccatg ggtaatgaca tgagattttc tttacaactt 4860  
tatactacga tgcattatca ttaccaccaat ttatgacccca taaccaaaaaa aaccataatg 4920

tgttaggttc attttcatt tttctaataa tttgcttcat gaatttttc tggagatatc 4980  
ttatctagat atttcttgcc aacatgttgc acctgataat tgatcgatt aatagttcag 5040  
aactttccaa aaactatgct gctcggtgtt ggctgtcatc catcagttt agaaaactat 5100  
tgacatgatt taagcctcggt cctgtactac taggaagggt aaactattgt tgcttccaaa 5160  
aatgtctttt aagggcgtgt tcagcaacaa tagttgtgt agtagctttt agctgttagt 5220  
tgtgctcgta gctgttagtg gttagtgtgt aactgttagc tgttcaagta gcggataaag 5280  
atattgtgt tcggtaaaag aagctgtcaa aatagctgtt tacaaagaat taataaaaaaa 5340  
ctcaaaacaaa gcttaatat ataatttatg caccactaaa gctaccccaa aagctacaaa 5400  
tttagcttt ttacaaacac tactaaaaca ctacttgtaa cactaaaagc tacttatact 5460  
actatttgc caaacattat tatttttct taattagtgt tttgacctag tcaagacact 5520  
aaaagctact tgaaaagctt ttgccgaaca cgccccttagt agacaagagg ggggaggggg 5580  
tcatcaagaa aatatgatta tactctcaac aaaaaaaaaa tgtaacttaa aaaaaataaaa 5640  
aataaataat tgactacttc aattaagaaa agaatagaat aaaaacatta cagtggatgt 5700  
ctcatccaca tccctaattt aatggcacaa tagaataatt gttttaaatt ttagaaatta 5760  
caacacaaga tgtaaattac tcttatcttc ctcttcgtaa tcttttact ctcccttac 5820  
ctcttccttt acctctacat aaaatagaga attagagatt gattaagata attataagat 5880  
tttagaaaca ttggtaaga aattctcaa caaacataat caagtaactc cattattta 5940  
gtttagtgac ttgctattt tcaccctaattt ttcaccatct accgcccctcc ttggacaata 6000  
ttgcccccttc cactttcttc actcttcctt cctcacgcat cttatcatct cttccacta 6060  
tcacctttaa aaaagtgtgt caggcacaac aaaaacgcctt ttatcaaccc acgcgaggcg 6120  
aagtacgtca ggcgcaacaa ggcgcgcacc taattctgtc ttttgcag gctgatggtg 6180  
cacttagttt taaaaagcgc agcaaagatg tgcctaggcg caaggcggtg aaaaaattgc 6240  
atccgtcagc agcggagtag aggctcacaa caataggtgc gaagaggcgt gcacgtacaa 6300  
aagaagcaaa aataagaaac tcaaataatga gacccagtgt ttaacatgta aattcgatac 6360  
ccagtgttta acatgttaat tcgattaaaa gcccttaattt aattgcatttga aattaattca 6420  
ttttaaccta tactaaaagc cctaataatgaa gaaaatccta gtttgcaggt tgaggaattt 6480  
ggaaaattga tgattgttgg atttgaaaaa attgttgcgg gcgatgaatg tgaggtggtt 6540  
tatggcacta gagaggttgg cgttcggtgc cgatgaagct ttccaagggtc attctctcct 6600  
tgtcttccttc ctatgccttag ctctcttccc tctccttaat ctctcttct tttctattct 6660  
ctctctttat cactacatta tgtttatttc tcgttcttcc cctatgtctt tcactggac 6720  
acttcggggg tatcttcatc ttttatctgc aatttgaagt ttgagaagct tccagagtcg 6780  
agtgttaaac ttttgcttct ttttttttaa tcttttgccc tttttctta gtggcccttg 6840

actagtgtatg	cacatgtgac	caattactaa	atgagctttt	attttgtctc	tcttcctttt	6900
caagcaattt	tttttaagta	aatcatctaa	aacaaagtac	tatccatttt	agttgtgtaa	6960
atggtgctat	tttaaaaaccg	cacaaaaatt	aaaaacataa	aaataaaggt	gtgcttcgca	7020
tacaagatgt	atgcgccttc	gtcttgcccc	ttttgagact	aagactacca	taagaactta	7080
gtcacttgag	aatggaatgg	gtgcaagatg	gacgacgata	attctaaaga	cctctagaag	7140
gatagtgtat	agtaactaat	acgaaccgaa	atataagttt	aactaaaatt	ttaaagtcta	7200
tatttccata	tggtatatgc	tggaatacac	gaaatgtcca	gaattttag	tggaccacga	7260
tccacacgtc	ttttcaggat	tctaggtgta	ttccaacgaa	aaatataaga	aaaccatatt	7320
ctactatctg	gttgttgtca	tccttcctt	gccggcgtga	cttctcatcc	ttttatTTT	7380
gtccggtgct	ggtgacacac	ttccttatga	tagtgtggg	caaagtaagg	tgtatgatatg	7440
gtgtttgtat	gaggtgtgg	gatTTTgtg	gtgggtgggt	gaagaggggt	ggttgcata	7500
agaaaggggt	aagagtcaat	gaggggtgga	aggggacaag	gggtatattt	gtaaatgcata	7560
gtaacattag	ggtgggtttt	agtaatTTT	gggaagttaa	tataaactac	cccctttgg	7620
tacaagagag	aataccgaa	ctactgctct	gatattttt	ttcacgttat	ttgatgtaat	7680
tacgcaatta	atttgtttc	tataagcttc	cgcacacaat	tgtgcata	aggctagtct	7740
aatatgagac	accaacataa	ctgactttct	tttgcacga	agttacctt	tcaagattt	7800
aacatagcat	caggattta	tttgttgtat	ctgtcatcct	tgttattgc	tttaattatg	7860
ctttgtatga	tgcattttac	cacttcgtat	gaaaaaaaa	gaaatttcat	tttagtgtca	7920
tttacatatt	acgagttgt	gacatgttt	aacatttgat	tttggaaatt	ttaagcctca	7980
tattatggag	atttatttga	cacaaatata	gccataattc	tccatcaact	tgtttctaga	8040
agtgttgctc	ttcctgatgt	acttgaattc	taatttagtt	ttatcagaca	ttatattata	8100
atgatatgtat	ttacaatTT	ttttagtgaa	aaaggattt	gagctcgcc	gaaggattgg	8160
ggcagagag	aggggatgg	aactaaacaa	cacagatgac	tcatttattt	aaggccaac	8220
aattgaattt	gctgttgatt	tcatctgtat	atactgctct	aggcttctat	tccaatgtaa	8280
tttataaaatc	caaggttagt	agcatgtttaa	gctttgtatt	cagtataatg	agacttata	8340
tttgcagttt	agatTTtagt	tgttgtatgt	gacttgtaaa	ttgttaactt	taagtgacgt	8400
cttgaggatt	atcgggacaa	tatactattt	ttttttcaa	a		8441

<210> 22  
<211> 462  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Sorghum bicolor - cDNA coding for CENH3

<400> 22  
atgagagtta aacacactgc tgccaggaaa tcaaccacca atggtcctcg ttcaaaggct 60  
cagaaatctc cgcgcatgtt gcaatcacca caatcgccctt ctagtagttc aaagcgcaaa 120  
tcacgcagaa acactgatgc aactcctcaa aagaagaagg ctaccgccc taagccggc 180  
actgtggcac tctggaaat acgcaaattt cagaagtcat tcaagccctt gattcctgct 240  
gccccttca ttcaaacagt gagagagatt actcaccagt ttgctcctta tgttggcgt 300  
tggcaagctg aagctctgat gccccttcaa gaggctgcag agaattttat tgtccgttt 360  
tttgaagatg gtatgcttg tgcaattcat gccaaacgag ttacactcat gaaaaaggat 420  
ttggagctcg cgcaaggat tggggcaga gagagggat gg 462

<210> 23  
<211> 154  
<212> PRT  
<213> Beta vulgaris

<400> 23

Met Arg Val Lys His Thr Ala Ala Arg Lys Ser Thr Thr Asn Gly Pro  
1 5 10 15

Arg Ser Lys Ala Gln Lys Ser Pro Arg Ser Leu Gln Ser Pro Gln Ser  
20 25 30

Pro Ser Ser Ser Ser Lys Arg Lys Ser Arg Arg Asn Thr Asp Ala Thr  
35 40 45

Pro Gln Lys Lys Ala Tyr Arg Arg Lys Pro Gly Thr Val Ala Leu  
50 55 60

Trp Glu Ile Arg Lys Phe Gln Lys Ser Phe Lys Pro Leu Ile Pro Ala  
65 70 75 80

Ala Pro Phe Ile Arg Thr Val Arg Glu Ile Thr His Gln Phe Ala Pro  
85 90 95

Tyr Val Gly Arg Trp Gln Ala Glu Ala Leu Met Ala Leu Gln Glu Ala  
100 105 110

Ala Glu Asn Phe Ile Val Arg Leu Phe Glu Asp Gly Met Leu Cys Ala  
115 120 125

Ile His Ala Lys Arg Val Thr Leu Met Lys Asp Leu Glu Leu Ala  
130 135 140

Arg Arg Ile Gly Gly Arg Glu Arg Gly Trp  
145 150

<210> 24  
 <211> 4597  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Z. mays CENH3-Mu-mutation  
  
 <220>  
 <221> misc\_recomb  
 <222> (676)..(685)  
 <223> transposon insertion  
  
 <400> 24

agagacggaa	agagtatata	tatttgttt	gctattattt	aaagtattaa	aagatagtgg	60
acgaatgaac	gtcctctatg	tttaaaaagaa	cgttttagag	gacgttgtgt	tgttgaagga	120
aatatgaaaa	aaaaatcttc	tgcataattt	gaagggagga	gcgtttacac	attacttcg	180
ggacctcaac	ccaaatatgt	caaggttgt	gagtggctca	gtgcggaaaa	aaaatcctat	240
atataccaga	tgtaaacact	atctttaca	gcctatcaca	ttcacattt	gaggttcaca	300
aagatagatc	aaaatttata	aaataatcat	ttaatatttt	ttttatttta	tttatatgga	360
taaggcagctg	gtgtatgtga	ggagctgtaa	aagatatttt	ttacatccga	gatgtaaaga	420
tttttttaa	ctcaatgctg	gttaccggct	gggaggacga	tgataaagaa	agcatctctc	480
actgcattcc	gggcccacta	ctcaaacgtt	cggcacgcca	ggttggcagg	tagccgttac	540
atcgataggc	actcggccac	tcgcacgcag	acaccacacc	agtgtgctca	gtgctcactg	600
ctcaccataa	taacgctgca	cctctttca	tttcaccatc	tcctgcccc	ttaaaaaaaaa	660
gactcaccgt	cgacacgccc	tcccgcccc	agagttctga	atcgaaaccg	tcggccacga	720
gagcagtgcg	aggcgcccac	cgcgatggct	cgaaccaagc	accaggccgt	gaggaagacg	780
gcggagaagc	ccaagaagaa	gctccagttc	gagcgctcag	gtaaccgggg	tcccgcgctc	840
ccccccgctt	cgcaaggcaga	cgctgtcgct	tctctccgac	cctggtgcta	agcacgttcc	900
ttgttccgtc	ttttgcaggt	ggtgcgagta	cctcggcgac	gccggtgagc	gcgtgcgtgc	960
ggggatcagt	tccctccctt	tgccttttt	tgttgggctg	ctcttacttg	cttgcaagct	1020
gtttgatgga	atgcagggaaa	gggctgctgg	gaccggggga	agagcggcgt	ctggaggtga	1080
ctcaggtgag	gacctattt	tcgttgctgg	atgctgggtt	tcgcttgcaa	tctaatttt	1140
ttgcaagatg	agggcgaatg	tgccagttcc	atgtgggtgt	catggtctcg	gagttactac	1200
cttaattgct	caccatagta	tgtttctta	aaaaaaacag	ttaagaagac	gaaaccacgc	1260
caccgctggc	ggccaggggac	tgtagcgctg	cgggagatca	ggaagtacca	gaagtccact	1320
gaaccgctca	tccccttgc	gccttcgtc	cgtgtggtgg	gtgcaggcgt	gtttgtcctc	1380
tgcatacgat	ggggttgttc	cgcattctgt	ctaatggaaa	gttattcttc	tgagaaaaaa	1440

aatgcaggtg agggagttaa ccaatttcgt aacaaacggg aaagtagagc gctataccgc 1500  
agaagccctc cttgcgctgc aagaggtcag ttatgaaaaa tgtcttatct ctctgttaag 1560  
atcctcttca tatacatagt tgctattgct atcgtgaagt ctttttttc tgttaattgg 1620  
tctggacta cttactagtc aggatttcat attgcgttt ttcctagtgg tgtgttagtta 1680  
aaaagtagtt taattgctt tagttaaaag gggtgttcag ggctaaagat caactatgag 1740  
aaaacagaaa ttttccaaat tcgatacccg acagcattat ggcctgcgc aatggaggtg 1800  
tttccggca aatactctag cctacctggg aagtaccttgc ggttgcgc tcatttcagg 1860  
aaagtaaaaaa ggaatgatct tcaacctcta atcgaaaaaa tcaacaacag gctggccttgc 1920  
ctggaaaggc aagatgttgt ccaaggctgg tatagaaaact ctgttaaaat cgatgctatc 1980  
cgcacaacca atctaccatc taatggttt tccaccta aatggctgc tgcaaacaat 2040  
tgacaaaata cgaagaaact tcctgtggag agggagcaat ccagaagttt gcagcggggg 2100  
tcactgcctc gtcaactggc ccgttaacttg cctccaaag aacaagggag gtcttggaat 2160  
tctgacacctt gatcgtttg cgagggggct aagactaaga tggctgtggc tacgatggaa 2220  
gagcaaagat agggcgtgga ctgccttgaa gtttccttgt gacaaaactg atgaagatct 2280  
cttcaatgct tccacaactg tcacggtagg caatggaaag atagctgaat tctggattc 2340  
tagttggatc caaggccaag cccctaagaa cattgcgcc acaactgttca agaaggaaaa 2400  
gaggaagaac atcacggcgc ccaaagcgc cactaacaac aattggattc gtttatgctc 2460  
accatacacg ggtgaggggg agtttagaga ggtcgtctct ctttggcagg ccataggtaa 2520  
catgcaagag cttaacgggtt tggaaagacaa catctcttg agatggacgg cagatggca 2580  
gtacagtgcgt acgagtgcac aaaaaatcca gttcgcatcc aatttcacta aaatgaaccc 2640  
ctgccctatt tggaaaggcta aagtggaaacc gaaatgccga tttttgctt ggacactact 2700  
tcataagaga attctgactg ccgataacct tcataaaaaga gtttgcaact cagcctcaga 2760  
aacaattccc cacttatgca aggattgccc cttagtaga gaggtgtgga acaaagttt 2820  
gtctcggcc aactttcctt tactgactgg gtctccagt gacacttctt tgtatgattg 2880  
gtggacggac atgtgcagcc tttgcagcag acaggcaaga agaggttcg acggctcg 2940  
atttcacttt tggtggact tatggctgga aagaaataac agaatcttc aaaggcagcg 3000  
tagaagtgtt gatcaagttt ctctggcagt caaggattat gcttagtagt gaagtcttagt 3060  
tggtttggac tagtggttt gttgctttc ttttaattt cttttagtt cttttatgt 3120  
tgtttcgtt tccttaagtt gctggagtc tgtattatcc tctttcttct aatatagatc 3180  
ggagcgcacaa acctttgccc cttcccttc aaaaaaaaaat taaaaggaa tttaactgct 3240  
ttccttagtgg tgttagttaaa atggatttca tattgcggcc tttccttagct tgcttgctat 3300  
tgattggact atagtgtatcc aaatgctgt aactttgtcg ctgtgttagg catggtaga 3360

gagcttagag tttgcattt ttcaataacct tgagactgca tttcatatac ataaattatt	3420
catgattatt tctttctct attgttctg gttaattaag agtttttaggt ttccatattt	3480
ttgtacgtgc atcattaaa ttcttgtatt gttttcggt cttgtctaca ggcagcagaa	3540
ttccacttga tagaactgtt tgaaatggcg aatctgtgtg ccatccatgc caagcgtgtc	3600
acaatcagta agttatcact gagtgaactc cttttctct gtagcattac tcctaattgaa	3660
tatgtgtat gcattttgggt tgcacgattc ttttagtgatt ctgcttcaga tggatatgat	3720
aaatcttagat gttatatttga agtggcgaat tgcttacgag cgaaaaatagt aatgttcaaa	3780
tagcgcaaag tgcaactgtt gacttttagt aggccattta tatggtttga ttaccaacaa	3840
atacgtcaat catatgattt gattatcaac aaaggaatca gctatatggg ttgattatca	3900
acaaaggaat cagcttaggtt tgcttatcaa cattcaacaa aggcataaag taatactcca	3960
tccgtttcaa tttataattt gtttgacttt ttttatctaa gtttgatcgg ctcgacttat	4020
aaaaaaaaat cataattatt gttaattttt gttgtgatat tggtagtat aatatacttt	4080
aatatgtgact ttgagttttt catttttcg caaaaaaaaaa tgaataggac gagccggtca	4140
aacgtgacac aaaaaagtca aacgaattat aatttggac acacggagta gtaaataatg	4200
taacaactta gagagtggga caaaaaaaaaa tctagtggtg ctaaatttag ttcagctttg	4260
tataaacaca agcattgatt gagaatctg acaactcaag gatctgttagg aaatgtgtta	4320
ccctaaatgt tttccttact gatgcagtgc aaaaggacat acaacttgca aggcgtatcg	4380
gaggaaggcg ttgggcatga tatataatat ccattctgat tgcatcattt ttgtgaattt	4440
gttttagga gctagacatt agtgttggt aatgctgcat gtttcctaatt cttttcgca	4500
gtctaacatc tgtggagttt gtatgttaca tggcaacagc tgaacatctg tggacttat	4560
ggcaacagcc gaagattgtg tctgtggat aactgg	4597

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Растение, обладающее биологической активностью гаплоидного индуктора и содержащее полинуклеотид, включающий нуклеотидную последовательность, кодирующую центромерный гистоновый Н3 (CENH3) белок, характеризующееся тем, что полинуклеотид содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности белка CENH3, при этом указанное изменение придает растению биологическую активность гаплоидного индуктора.
2. Растение по п. 1, отличающееся тем, что по меньшей мере одна мутация, вызывающая изменение, является точечной мутацией, инсерцией или делецией, либо мутацией в сайте сплайсинга, и/или изменение представляет собой замену одной или нескольких аминокислот, инсерцию одной или нескольких аминокислот, либо делецию одной или нескольких аминокислот.
3. Растение по п. 1 или п. 2, отличающееся тем, что по меньшей мере одна мутация вызывает изменение аминокислотной последовательности белка CENH3 в по меньшей мере одном сегменте аминокислотной последовательности белка CENH3, выбираемом из группы, включающей:
  - a) N-терминальный хвостовой домен, соответствующий нуклеотидам в положении 1-246 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11,
  - b)  $\alpha$ N-спираль, соответствующую нуклеотидам в положении 247-291 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11,
  - c)  $\alpha$ 1-спираль, соответствующую нуклеотидам в положении 307-339 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11,
  - d) петлю1, соответствующую нуклеотидам в положении 340-378 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11,
  - e)  $\alpha$ 2-спираль, соответствующую нуклеотидам в положении 379-465 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11,
  - f) петлю2, соответствующую нуклеотидам в положении 466-486 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11,
  - g)  $\alpha$ 3-спираль, соответствующую нуклеотидам в положении 487-516 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11, либо

- h) С-терминальный домен, соответствующий нуклеотидам в положении 517-534 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana*, приведенного в SEQ ID NO. 11.
4. Растение по любому из предшествующих пунктов, отличающееся тем, что по меньшей мере одна мутация вызывает аминокислотную замену или делецию специфической или неспецифической аминокислоты, представленной в Таблице 1, Таблице 2, Таблице 3, Таблице 4, Таблице 5, Таблице 6, Таблице 7 или Таблице 8, либо аминокислотную замену или делецию специфической или неспецифической аминокислоты SEQ ID NO. 1, SEQ ID NO. 2, SEQ ID NO. 3, SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 7, SEQ ID NO. 8 или SEQ ID NO. 9.
5. Растение по любому из предшествующих пунктов, отличающееся тем, что по меньшей мере одна мутация вызывает аминокислотную замену специфической или неспецифической аминокислоты в сегменте, выбираемом из группы, включающей:
- N-терминальный хвостовой домен, где
    - аминокислотный аргинин в положении 3 последовательности SEQ ID NO. 1 заменен предпочтительно на лизин, либо
    - аминокислотный аргинин в положении 2 последовательности SEQ ID NO. 23 заменен предпочтительно на лизин, либо
    - аминокислотный аргинин в положении 10 последовательности SEQ ID NO. 1 заменен предпочтительно на фенилаланин, либо
    - аминокислотный серин в положении 9 последовательности SEQ ID NO. 14 заменен предпочтительно на фенилаланин, либо
    - аминокислотный аргинин в положении 16 последовательности SEQ ID NO. 14 заменен предпочтительно на глутамин, либо
    - аминокислотный серин в положении 24 последовательности SEQ ID NO. 14 заменен предпочтительно на лейцин, либо
    - аминокислотный серин в положении 24 последовательности SEQ ID NO. 14 заменен предпочтительно на лейцин, либо
    - аминокислотный аланин в положении 25 последовательности SEQ ID NO. 17 заменен предпочтительно на треонин, либо
    - аминокислотная глутаминовая кислота в положении 29 последовательности SEQ ID NO. 14 заменена предпочтительно на лизин, либо
    - аминокислотный глицин в положении 30 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагиновую кислоту, либо
    - аминокислотный аланин в положении 33 последовательности SEQ ID NO. 14 заменен предпочтительно на треонин, либо

- xii) аминокислотный пролин в положении 35 последовательности SEQ ID NO. 14 заменен предпочтительно на лейцин, либо
- xiii) аминокислотная глутаминовая кислота в положении 35 последовательности SEQ ID NO. 20 заменена предпочтительно на лизин, либо
- xiv) аминокислотный серин в положении 41 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагин, либо
- xv) аминокислотный глицин в положении 43 последовательности SEQ ID NO. 14 заменен предпочтительно на глутаминовую кислоту, либо
- xvi) аминокислотный пролин в положении 50 последовательности SEQ ID NO. 14 заменен предпочтительно на серин, либо
- xvii) аминокислотный пролин в положении 55 последовательности SEQ ID NO. 14 заменен предпочтительно на лейцин, либо
- xviii) аминокислотный глицин в положении 57 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагиновую кислоту, либо
- xix) аминокислотный глицин в положении 61 последовательности SEQ ID NO. 14 заменен предпочтительно на глутаминовую кислоту, либо
- xx) аминокислотный аргинин в положении 65 последовательности SEQ ID NO. 14 заменен предпочтительно на глутамин, либо
- xxi) аминокислотный аргинин в положении 65 последовательности SEQ ID NO. 14 заменен предпочтительно на стоп-сигнал, либо
- xxii) аминокислотный пролин в положении 71 последовательности SEQ ID NO. 14 заменен предпочтительно на серин, либо
- xxiii) аминокислотная аспарагиновая кислота в положении 46 последовательности SEQ ID NO. 23 заменена предпочтительно на аспарагин или глицин, либо
- xxiv) аминокислотный лизин в положении 71 последовательности SEQ ID NO. 2 заменен предпочтительно на серин, либо
- xxv) аминокислотный пролин в положении 56 последовательности SEQ ID NO. 20 заменен предпочтительно на серин, либо
- xxvi) аминокислотный пролин в положении 14 последовательности SEQ ID NO. 2 заменен предпочтительно на валин, либо
- xxvii) аминокислотный аланин в положении 62 последовательности SEQ ID NO. 17 заменен предпочтительно на валин, либо

b)  $\alpha$ N-спираль, где

- i) аминокислотный треонин в положении 2 последовательности SEQ ID NO. 3 заменен предпочтительно на серин, либо

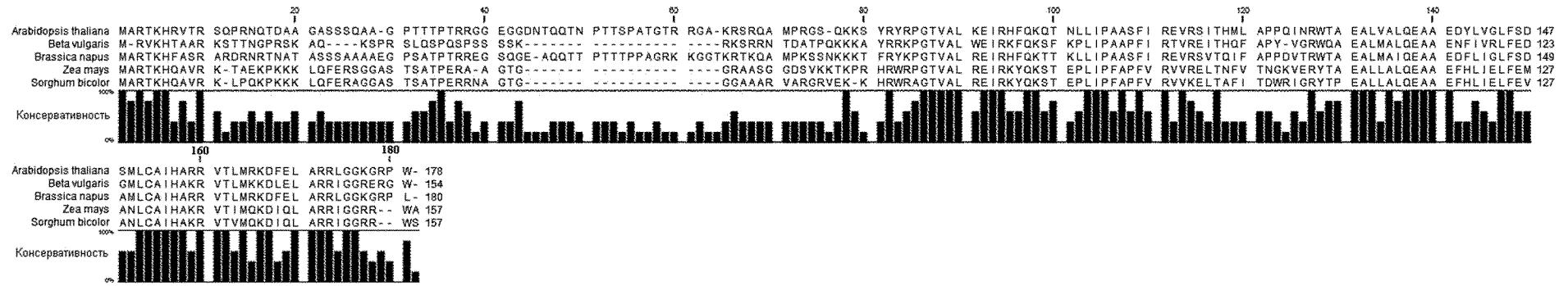
- ii) аминокислотный треонин в положении 64 последовательности SEQ ID NO.17 заменен предпочтительно на серин, либо
  - c)  $\alpha 1$ -спираль, где
    - i) аминокислотный аланин в положении 1 последовательности SEQ ID NO. 4 заменен предпочтительно на треонин, либо
    - ii) аминокислотный аланин в положении 105 последовательности SEQ ID NO. 14 заменен предпочтительно на треонин, либо
    - iii) аминокислотный аргинин в положении 6 последовательности SEQ ID NO. 4 заменен предпочтительно на глутамин, либо
    - iv) аминокислотный аргинин в положении 110 последовательности SEQ ID NO. 14 заменен предпочтительно на глутамин, либо
    - v) аминокислотный валин в положении 89 последовательности SEQ ID NO. 20 заменен предпочтительно на метионин, либо
    - vi) аминокислотная глутаминовая кислота в положении 10 последовательности SEQ ID NO. 4 заменена предпочтительно на аспарагин, либо
    - vii) аминокислотный серин в положении 114 последовательности SEQ ID NO. 14 заменен предпочтительно на аспарагин, либо
- d) петлю1, где
  - i) аминокислотный аспарагин в положении 2 последовательности SEQ ID NO. 5 заменен предпочтительно на валин, либо
  - ii) аминокислотный аланин в положении 95 последовательности SEQ ID NO. 17 заменен предпочтительно на валин, либо
  - iii) аминокислотный пролин в положении 6 последовательности SEQ ID NO. 5 заменен предпочтительно на серин, либо
  - iv) аминокислотный пролин в положении 121 последовательности SEQ ID NO. 14 заменен предпочтительно на серин, либо
  - v) аминокислотный триптофан в положении 12 последовательности SEQ ID NO. 5 заменен предпочтительно на стоп-сигнал, либо
  - vi) аминокислотный триптофан в положении 127 последовательности SEQ ID NO. 4 заменен предпочтительно на стоп-сигнал, либо
- e)  $\alpha 2$ -спираль, где
  - i) аминокислотный аланин в положении 1 последовательности SEQ ID NO. 6 заменен предпочтительно на треонин, либо
  - ii) аминокислотный аланин в положении 107 последовательности SEQ ID NO. 20 заменен предпочтительно на треонин, либо
  - iii) аминокислотный лейцин в положении 4 последовательности SEQ ID NO. 6 заменен предпочтительно на фенилаланин или глутамин, либо

- iv) аминокислотный лейцин в положении 132 последовательности SEQ ID NO. 14 заменен предпочтительно на фенилаланин или глутамин, либо
  - v) аминокислотный лейцин в положении 7 последовательности SEQ ID NO. 6 заменен предпочтительно на пролин, либо
  - vi) vi. аминокислотный лейцин в положении 109 последовательности SEQ ID NO. 23 заменен предпочтительно на лейцин, либо
  - vii) аминокислотный глутамин в положении 8 последовательности SEQ ID NO. 6 заменен предпочтительно на стоп-сигнал или лейцин, либо
  - viii) аминокислотный глутамин в положении 114 последовательности SEQ ID NO. 20 заменен предпочтительно на стоп-сигнал или лейцин, либо
  - ix) аминокислотный аланин в положении 10 последовательности SEQ ID NO. 6 заменен предпочтительно на треонин, либо
  - x) x. аминокислотный аланин в положении 138 последовательности SEQ ID NO. 14 заменен предпочтительно на треонин, либо
  - xi) аминокислотный цистеин в положении 25 последовательности SEQ ID NO. 6 заменен предпочтительно на тирозин, либо
  - xii) аминокислотный цистеин в положении 153 последовательности SEQ ID NO. 14 заменен предпочтительно на тирозин, либо
  - xiii) аминокислотный аланин в положении 26 последовательности SEQ ID NO. 6 заменен предпочтительно на валин, либо
  - xiv) аминокислотный аланин в положении 154 последовательности SEQ ID NO. 14 заменен предпочтительно на валин, либо
- f) петлю2, где
- i) аминокислотный аргинин в положении 2 последовательности SEQ ID NO. 7 заменен предпочтительно на гистидин, либо
  - ii) аминокислотный аргинин в положении 159 последовательности SEQ ID NO. 14 заменен предпочтительно на гистидин, либо
  - iii) аминокислотный валин в положении 3 последовательности SEQ ID NO. 7 заменен предпочтительно на изолейцин, либо
  - iv) аминокислотный валин в положении 160 последовательности SEQ ID NO. 14 заменен предпочтительно на изолейцин, либо
  - v) аминокислотный треонин в положении 4 последовательности SEQ ID NO. 7 заменен предпочтительно на изолейцин, либо
  - vi) аминокислотный треонин в положении 139 последовательности SEQ ID NO. 20 заменен предпочтительно на изолейцин, либо
- g) α3-спираль, где

- i) аминокислотная аспарагиновая кислота в положении 2 последовательности SEQ ID NO. 8 заменена предпочтительно на аспарагин, либо
  - ii) аминокислотная аспарагиновая кислота в положении 166 последовательности SEQ ID NO. 14 заменена предпочтительно на аспарагин, либо
  - iii) аминокислотная глутаминовая кислота в положении 4 последовательности SEQ ID NO. 8 заменена предпочтительно на лизин, либо
  - iv) аминокислотная глутаминовая кислота в положении 168 последовательности SEQ ID NO. 14 заменена предпочтительно на лизин, либо
  - v) аминокислотный аргинин в положении 8 последовательности SEQ ID NO. 8 заменен предпочтительно на гистидин, либо
  - vi) аминокислотный аргинин в положении 172 последовательности SEQ ID NO. 14 заменен предпочтительно на гистидин, либо
  - vii) аминокислотный лейцин в положении 9 последовательности SEQ ID NO. 8 заменен предпочтительно на фенилаланин, либо
  - viii) аминокислотный лейцин в положении 173 последовательности SEQ ID NO. 14 заменен предпочтительно на фенилаланин, либо
  - ix) аминокислотный глицин в положении 10 последовательности SEQ ID NO. 8 заменен предпочтительно на глутаминовую кислоту, либо
  - x) аминокислотный глицин в положении 174 последовательности SEQ ID NO. 14 заменен предпочтительно на глутаминовую кислоту, либо
- h) С-терминальный домен, где
- i) аминокислотный глицин в положении 3 последовательности SEQ ID NO. 9 заменен предпочтительно на гистидин, либо
  - ii) аминокислотный аргинин в положении 155 последовательности SEQ ID NO. 20 заменен предпочтительно на гистидин, либо
  - iii) аминокислотный аргинин в положении 4 последовательности SEQ ID NO. 9 заменен предпочтительно на лизин, либо
  - iv) аминокислотный аргинин в положении 178 последовательности SEQ ID NO. 14 заменен предпочтительно на лизин, либо
  - v) аминокислотный серин в положении 157 последовательности SEQ ID NO. 17 заменен предпочтительно на лейцин.
6. Растение по любому из предшествующих пунктов, отличающееся тем, что скрещивание с растением дикого типа или растением, экспрессирующим белок CENH3 дикого типа, дает по меньшей мере 0,1% гаплоидного потомства.

7. Растение по любому из предшествующих пунктов, отличающееся тем, что полинуклеотид, содержащий по меньшей мере одну мутацию, представляет собой эндогенный ген или трансген.
8. Часть растения по любому из предшествующих пунктов, характеризующаяся тем, что она предпочтительно представляет собой вегетативный орган побега, корень, цветок или цветковый орган, семя, плод, семязачаток, зародыш, растительную ткань или растительную клетку.
9. Способ создания гаплоидного растения, включающий стадии:
  - a. скрещивание растения по пп. 1-7 с растением, экспрессирующим белок CENH3 дикого типа, и
  - b. идентификацию гаплоидного потомства растения, получаемого на стадии скрещивания.
10. Способ создания удвоенного гаплоидного растения, включающий стадии:
  - a. скрещивания растения по пп. 1-7 с растением, экспрессирующим белок CENH3 дикого типа,
  - b. идентификацию гаплоидного потомства растения, получаемого на стадии скрещивания, и
  - c. перевод гаплоидного растения в удвоенный гаплоид, предпочтительно обработкой колхицином или посредством спонтанного удвоения хромосом.
11. Способ облегчения цитоплазматического обмена, включающий стадии:
  - x. скрещивания растения по пп. 1-7 в качестве материнского родителя с растением, экспрессирующим белок CENH3 дикого типа, в качестве отцовского родителя,
  - y. получения гаплоидного потомства растения, содержащего хромосомы отцовского родителя и цитоплазму материнского родителя.
12. Способ создания растения по пп. 1-7, включающий стадии:
  - i. обработки семян растения достаточным количеством мутагена, предпочтительно этилметансульфоната, для получения растений M1,
  - ii. обеспечения производства достаточного количества фертильных растений M2,
  - iii. выделения геномной ДНК из растений M2 и
  - iv. отбора индивидуумов, содержащих по меньшей мере одну мутацию в полинуклеотиде, включающем нуклеотидную последовательность, кодирующую центромерный гистоновый H3 (CENH3) белок, где по меньшей мере одна мутация вызывает изменение аминокислотной последовательности белка CENH3 в по меньшей мере сегменте аминокислотной последовательности белка CENH3, выбиремом из группы, включающей:

- a. N-терминальный хвостовой домен, соответствующий нуклеотидам в положении 1-246 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11,
  - b.  $\alpha$ N-спираль, соответствующую нуклеотидам в положении 247-291 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11,
  - c.  $\alpha$ 1-спираль, соответствующую нуклеотидам в положении 307-339 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11,
  - d. петлю1, соответствующую нуклеотидам в положении 340-378 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11,
  - e.  $\alpha$ 2-спираль, соответствующую нуклеотидам в положении 379-465 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11,
  - f. петлю2, соответствующую нуклеотидам в положении 466-486 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11,
  - g.  $\alpha$ 3-спираль, соответствующую нуклеотидам в положении 487-516 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11, либо
  - h. С-терминальный домен, соответствующий нуклеотидам в положении 517-534 последовательности SEQ ID NO. 10 белка CENH3 из *Arabidopsis thaliana* последовательности SEQ ID NO. 11.
13. Полинуклеотид, включающий нуклеотидную последовательность, кодирующую по меньшей мере один сегмент аминокислотной последовательности, кодирующей CENH3, характеризующийся тем, что он содержит по меньшей мере одну мутацию, вызывающую изменение аминокислотной последовательности по меньшей мере одного сегмента.
14. Вектор, содержащий полинуклеотид по п. 13.
15. Растительная клетка или клетка-хозяин, содержащая полинуклеотид по п. 13 или вектор по п. 14 в качестве трансгена.
16. Способ создания растения по пп.1-7, включающий стадии:
- уу) трансформации растительной клетки, содержащей полинуклеотид по п. 13 или вектор по п.14 и
  - хх) регенерации растения, обладающего биологической активностью гаплоидного индуктора, из растительной клетки.



Фиг. 1