Identification and molecular characterization of ethylene-insensitive mutants in *Cucurbita pepo*

Identificación y caracterización molecular de mutantes insensibles a etileno en *Cucurbita pepo*

Alicia García Fuentes

Almería, 26 de noviembre de 2019

Ph.D. DISSERTATION

Identification and molecular characterization of ethylene-insensitive mutants in *Cucurbita pepo*

Área de Genética
Departamento de Biología y Geología

Escuela Superior de Ingeniería
Universidad de Almería

Fdo. Alicia García Fuentes
Almería, 26 de noviembre de 2019

Dpto. de Biología y Geología

Área de Genética

Universidad de Almería

Dr. Manuel Jamilena Quesada, Catedrático de Genética de la Universidad de Almería,

HACE CONSTAR:

Que el presente trabajo se ha realizado bajo mi dirección y recoge la labor realizada por la licenciada Alicia García Fuentes para optar al Grado de Doctor en Biología.

Almería, 26 de noviembre de 2019

Fdo. Manuel Jamilena Quesada

A mis padres,
a mis abuelas.

AGRADECIMIENTOS

Parece que fue ayer cuando decidí comenzar esta etapa de mi carrera profesional, con bastante incertidumbre al inicio, pero contenta y orgullosa del aprendizaje y cada uno de los logros conseguidos a lo largo de estos años. En estas líneas me gustaría agradecer a las personas que me han acompañado durante el transcurso de mi tesis doctoral y que me han hecho crecer tanto en lo profesional, como en lo personal.

En primer lugar, quiero dar las gracias a mi director, Manuel Jamilena, por haber confiado en mí desde el primer momento para la realización de este proyecto. Sin tu apoyo y ayuda, sin duda, no habría llegado hasta aquí. Gracias por haberme guiado durante este camino y, sobre todo, por transmitirme la ilusión necesaria para la consecución de cada uno de los objetivos que nos hemos propuesto. Nos hemos adentrado en experimentos totalmente desconocidos y has puesto toda la confianza en mí demostrando que el esfuerzo, la responsabilidad y la constancia son fundamentales para obtener cada uno de los resultados de esta tesis doctoral. Gracias.

Al resto de profesores con los que he tenido el gusto de compartir estos años, Juan Luis Valenzuela, María del Mar Rebolloso y Dolores Garrido, gracias por los consejos, los ánimos y la ayuda.

A cada uno de los compañeros del laboratorio que de una forma u otra habéis formado parte de esta etapa. Recuerdo mis inicios en el laboratorio con Susana, Cecilia y Zoraida, cada uno de vuestros consejos me sirvieron para comenzar esta andadura. A Zoraida por los ratos de charla necesarios, gracias por haber estado ahí. A la reincorporación de Cecilia con nuestros largos debates y risas, mil gracias por todo el apoyo en este último año, conseguirás cualquier meta que te propongas porque te lo mereces. A los compañeros actuales Jonathan, Gustavo y Jessica, desearos todo el ánimo y suerte para completar vuestras tesis. A todos los alumnos que han colaborado en el desarrollo de los experimentos durante estos años, gracias por vuestra atención y dedicación.

Gracias a Encarni, porque es muy difícil encontrar a personas con la profesionalidad y la personalidad que te caracterizan. Por tu ayuda incondicional en todo momento, por nuestras horas en el invernadero fenotipando y su típica frase… venga que ya falta poco…, por las tardes interminables delante de la ‘computadora’ con un programilla llamado R o pantallas en negro llenas de comandos, y de nuevo su frase… venga que esto tiene que salir… Qué decir, has sido clave para que yo pueda estar escribiendo estas líneas. Estoy muy contenta de haber encontrado una persona como tú y me llevo una amiga para toda la vida, estoy segura de que conseguirás lo que te propongas.

Gracias por la oportunidad de asistir a cada uno de los congresos, por la oportunidad de conocer a gente que ha hecho que me apasione un poquito más lo que hacemos y empujarme a seguir trabajando para lograr mi objetivo.

Gracias al Prof. Giovannoni, por haberme acogido en su laboratorio durante mi estancia en Ithaca. Sólo tengo palabras de agradecimiento por toda la enseñanza y ayuda recibida de un gran profesional. Sin duda, fue una gran oportunidad poder estar en un centro como BTI. A Antonio, Betsy, Teo, Dann, Yimin, Xuedong, Julia, Itay, Ari… me he podido enriquecer de cada uno de vosotros durante esta gran experiencia. Y cómo no a mis colombianos Camila y Francisco, qué decir de los tíos, me quedo con cada uno de los momentos juntos en los que

nunca faltaba una risa, como vosotros decís, os extraño. Esta etapa americana fue todo un reto que ha transcendido tanto lo profesional como lo personal.

A los de toda la vida, a los que han hecho tantas versiones de lo que hago que han llegado a la conclusión de que estudio 'curcutáceos'. A Irene, Analy, Lucía, Neiva, Mónica, Marina, Álvaro, Vicente y, a las últimas incorporaciones, Alberto, Kiko y Aitor, gracias por estar ahí siempre, en los buenos y en los malos momentos. Porque me habéis acompañado en esta etapa y estoy segura de que lo haréis en todas las aventuras que nos quedan por vivir. A Raúl que, aunque te has incorporado en la última etapa, me has apoyado incondicionalmente a pesar de todas las dificultades. Gracias.

A María, por tu gran ayuda, por sacar una risa incluso cuando estábamos en uno de nuestros bucles, no podías faltar en estos agradecimientos por haber sido una gran compañera y amiga. Lucha por tu objetivo porque lo conseguirás.

A las peques del máster, bendita coincidencia que me hizo conocer a dos de las personas que más han marcado esta etapa. María y Cristina, por muchos más viernes de, ¿niñas que hacéis? y pozos que acaban en risas. Gracias por estar ahí siempre.

Gracias a las personas que formaron parte de mi etapa granadina y que siguen ahí. A Antonio o Mariolilla porque cada vez que nos vemos parece que se ha parado el tiempo. A mis geniles, que con nuestras locuras y cada una de nuestras reuniones llenas de risas y buenos recuerdos habéis hecho el camino más fácil, estoy feliz de teneros siempre cerca. A mis pezuñitas, porque aquella gran etapa italiana que nos unió me sirvió para descubriros y alegrarme cada día más. A Marilina, que cada reencuentro, sea donde sea, nos llena de alegría. A Rubén que, aunque pasen meses, siempre hay lugar para un interminable café. Es difícil encontrar a una persona que le apasione tanto lo que hace como a ti. Estoy segura de que llegarás muy lejos.

Gracias sobre todo a mi familia, a mis padres Rosario y José Antonio. por educarme como lo habéis hecho y darme vuestro apoyo incluso sin saber muy bien dónde me metía. Vosotros sois los responsables de que esté escribiendo estas líneas, por habernos enseñado que el sacrificio y la responsabilidad son las piezas clave para conseguir cada uno de nuestros objetivos y que, finalmente, cada una de nosotras estemos donde queramos estar. En especial a mi madre por la ayuda incondicional recibida y, como no, por el manejo del Photoshop, que sin ella la calidad de las imágenes de esta tesis dejarían mucho que desear. A mis hermanas Inma y Evelin, porque sin ellas no habría sido lo mismo y cada una, con vuestra forma de ser, me habéis apoyado para que consiga llegar hasta aquí, estando siempre a mi lado. A Fran, que ya es uno más.

Gracias al resto de mi familia, no puedo escribir todos vuestros nombres porque, afortunadamente, llenaría una página entera. Gracias por toda la alegría y el apoyo que me habéis dado siempre, no puedo estar más orgullosa de la familia que tengo. Este año hemos obtenido una lección de fortaleza y valentía, llevándonos un palo muy grande, y por eso tenemos que seguir así, luchando, porque siempre lo hemos hecho estando unidos. A la pequeña Alba por sacarnos siempre una sonrisa y por las nuevas incorporaciones que vendrán. A mi abuela Paca que ha demostrado ser el pilar fundamental y ser la persona más especial.

Y para finalizar, a todas aquellas personas que, aunque no os haya mencionado, habéis formado parte de este camino. GRACIAS.

Scientific contributions directly related to the doctoral thesis

Peer-Reviewed Articles

García A., Aguado E., Parra G., Manzano S., Martínez C., Megías Z., et al. (2018). Phenomic and genomic characterization of a mutant platform in *Cucurbita pepo*. Frontiers in Plant Science. 9, 1049. doi:10.3389/fpls.2018.01049.

García A., Aguado E., Martínez C., Loska D., Beltrán S., Valenzuela J. L., et al. (2019). The ethylene receptors *CpETR1A* and *CpETR2B* cooperate in the control of sex determination in *Cucurbita pepo*. Journal of Experimental Botany. doi:10.1093/jxb/erz417.

García A., Aguado E., Martínez C., Garrido D., Jamilena M. (2019). Two androecious mutations reveal the crucial role of ethylene receptors in the initiation of female flower development in *Cucurbita pepo.* Submitted to Plant Journal.

García A., Manzano S., Martínez C., Megías Z., Gázquez, J. C., Jamilena M. (2017). Isolation and characterisation of ethylene insensitive mutants from a collection of *Cucurbita pepo* L. Acta Horticulturae. 1151, 151–156. doi:10.17660/ActaHortic.2017.1151.24.

García A., Manzano S., Megías Z., Aguado E., Martinez C., Garrido D., et al. (2018). Use of mutant platforms to discover novel postharvest fruit-quality traits in *Cucurbita pepo*. Acta Horticulturae. 367–373. doi:10.17660/ActaHortic.2018.1194.53.

García A., Valenzuela J. L., Manzano S., Cebrián G., Romero J., Aguado E., et al. (2019). Postharvest fruit quality in ethylene insensitive mutants of zucchini squash. Acta Horticulturae. 217–222. doi:10.17660/ActaHortic.2019.1256.30.

Proceedings and conference communications

García A., Manzano S., Aguado E., Megías Z., Martínez C., Garrido D., et al. (2016). Isolation and characterization of three recessive andromonoecious mutants of *Cucurbita pepo*. Cucurbitaceae, XIth Eucarpia Meeting on Cucurbit Genetics and Breeding, July 24-28, 2016, Warsaw, Poland, July 24-28. 45–49. ISBN: 978-83-7987-896-3.

García A., Aguado E., Megías Z., Manzano S., Martínez C., Rebolloso M. M., Garrido D., Valenzuela J. L., Jamilena M. (2016). Comportamiento postcosecha de tres mutantes insensibles a etileno en calabacín (*Cucurbita pepo* L.). IX Simposio Ibérico de Maduración y Poscosecha, Lisboa, Portugal, Nov 2-4 2016. ISBN: 978-972-8936-25-9.

García A., Manzano S., Martínez C., Megías Z., Aguado E., Cebrián G., Romero J., Loska D., Beltran S., Cañizares J., Garrido D., Jamilena M. (2017). Evaluation of an EMS-induced squash library using phenomic and genomic approaches. The XIV Solanaceae and III Cucurbitaceae Genomics Joint Conference, Valencia, Spain, Sept 3-6, 2017.

García A., Romero J., Cebrián G., Aguado E., Manzano S., Guzmán M., Garrido D., Jamilena M. (2017) Rapid high-throughput phenotyping of an EMS mutant platform of *Cucurbita pepo* for tolerance to abiotic and biotic stresses. The COST/EPPN2020 workshop "Current and future applications of phenotyping for plant breeding". Novi Sad, Serbia, Sept 29-30, 2017.

García A., Megías Z., Aguado E., Manzano S., Rosales R., Garrido D., Valenzuela J. L., Jamilena M. (2018). Cambios transcriptómicos en respuesta a tratamientos postcosecha que mejoran la tolerancia al frío en calabacín. XII Simposio Nacional y X Ibérico sobre maduración y postcosecha, Badajoz, España, Jun 4-7, 2018.

García A., Aguado E., Manzano M., Romero R., Cebrián G., Garrido D., Jamilena M. (2018). Genetic interaction between *EIN1*, *EIN2* and *EIN3* in the regulation of sex expression and sex determination in squash. Cucurbitaceae, Davis, California, USA, Nov 12-15, 2018.

García A., Aguado E., Romero J., Cebrián G., Martínez C., Garrido D., Jamilena M. (2019). Characterization of two novel androecious single mutants in *Cucurbita pepo* L. VI International Symposium on Cucurbits, Gent, Belgium, Jun 30-July 4, 2019.

Other contributions

Other associated Peer-Reviewed Articles

Megías Z., Martínez C., Manzano S., **García A.**, Rebolloso-Fuentes M., Garrido D., et al. (2015). Individual shrink wrapping of zucchini fruit improves postharvest chilling tolerance associated with a reduction in ethylene production and oxidative stress metabolites. PLoS One 10, e0133058. doi:10.1371/journal.pone.0133058.

Megías Z., Martínez C., Manzano S., **García, A.**, Rebolloso-Fuentes M. M, Valenzuela J. L., et al. (2016). Ethylene biosynthesis and signaling elements involved in chilling injury and other postharvest quality traits in the non-climacteric fruit of zucchini (*Cucurbita pepo*). Postharvest Biology and Technology, 113, 48–57. doi:http://dx.doi.org/10.1016/j.postharvbio.2015.11.001.

Manzano S., Aguado E., Martínez C., Megías Z., **García A.**, Jamilena, M. (2016). The ethylene biosynthesis gene *CitACS4* regulates monoecy/andromonoecy in watermelon (*Citrullus lanatus*). PLoS One, 11, e0154362. doi:10.1371/journal.pone.0154362.

Megías Z., Manzano S., Martínez C., **García A.**, Aguado E., Garrido D., et al. (2017). Postharvest cold tolerance in summer squash and its association with reduced cold-induced ethylene production. Euphytica, 213. doi:10.1007/s10681-016-1805-0.

Manzano S., Megías Z., Martínez C., **García A.**, Aguado E., Chileh T., et al. (2017). Overexpression of a flower-specific aerolysin-like protein from the dioecious plant *Rumex acetosa* alters flower development and induces male sterility in transgenic tobacco. Plant Journal, 89, 58–72. doi:10.1111/tpj.13322.

Aguado E., **García A.**, Manzano S., Valenzuela J. L., Cuevas J., Pinillos V., et al. (2018). The sex-determining gene *CitACS4* is a pleiotropic regulator of flower and fruit development in watermelon (*Citrullus lanatus*). Plant Reproduction, 1–16. doi:10.1007/s00497-018-0346-1.

Martínez C., Manzano S., Megías Z., **García A.**, Garrido R., Paris H. S., et al. (2017). Screening of *Cucurbita* germplasm for ToLCNDV resistance under natural greenhouse conditions. Acta Horticulturae, 1151, 57–62. doi:10.17660/ActaHortic.2017.1151.10.

Megias Z., González-Rodríguez L.J., Aguado E., **García A.**, Manzano S., Rebolloso M.M., Valenzuela J.L., Jamilena M. (2018). Effect of cold storage time on chilling injury in two zucchini cultivars. Acta Horticulturae, 1194, pp. 479-485. doi: 10.17660/ActaHortic.2018.1194.70

Megias Z., Manzano S., Martinez C., **García A.**, Aguado E., Garrido D., et al. (2018). Breeding for postharvest cold tolerance in zucchini squash. Acta Horticulturae, 357–362. doi:10.17660/ActaHortic.2018.1194.51.

Resumen

El genoma de *Cucurbita pepo* comprende 263 Mb y 34,240 genes organizados en 20 cromosomas diferentes. Para mejorar nuestra comprensión de la función génica, hemos generado una plataforma de mutantes EMS que consta de 3.751 familias M_2 independientes. La calidad de la colección se ha evaluado en base a los resultados de fenotipado y re-secuenciación del genoma completo (WGS). La evaluación fenotípica de toda la plataforma en la etapa de plántula ha demostrado que la tasa de variación para rasgos fácilmente observables es de más del 10%. El porcentaje de familias con plántulas albinas o cloróticas excedió el 3%, similar o mayor a lo encontrado en otras colecciones EMS de cultivos de Cucurbitáceas. Tras evaluar 4 plantas adultas de 300 familias independientes, se encontró que más del 28% de las familias tenían un efecto fenotípico observable sobre diferentes rasgos vegetativos y reproductivos, incluyendo: el vigor de las plantas; el tamaño y la forma de las hojas; la expresión sexual y la determinación del sexo; y la producción y desarrollo de frutos. Dos muestras de ADN genómico, derivadas de 20 plantas de dos familias mutantes, se sometieron a WGS utilizando la metodología NGS con el fin de estimar la densidad, el espectro, la distribución y el impacto de las mutaciones inducidas por EMS. El número de mutaciones EMS en los genomas de las familias L1 y L2 fue de 1.704 y 859, respectivamente, lo que representa una densidad de 11,8 y 6 mutaciones por Mb, respectivamente. Como se esperaba, las mutaciones predominantes inducidas por EMS fueron las transiciones C>T y G>A (80,3% en L1 y 61% en L2), que se encontraron distribuidas al azar a lo largo de los 20 cromosomas de *C. pepo*. Las mutaciones afectaron principalmente a las regiones intergénicas, pero el 7,9 y el 6% de las mutaciones EMS identificadas en L1 y L2, respectivamente, se ubicaron en el exoma, y el 0,4 y el 0,2% podían tener un impacto moderado o alto en la función génica. Estos resultados han proporcionado información sobre el uso potencial de la plataforma mutante en el descubrimiento de nuevos alelos tanto para la genómica funcional como para la mejora genética de las especies del género *Cucurbita*.

El etileno es el regulador clave de la determinación del sexo en las especies monoicas de la familia *Cucurbitaceae*. Esta hormona regula la transición floral femenina y la conversión de cada meristemo floral en una flor femenina o masculina, lo que se logra deteniendo el crecimiento del estambre o el carpelo en las primeras etapas del desarrollo de la flor. Muchos de los genes reguladores descubiertos hasta la fecha codifican para enzimas de biosíntesis de etileno, pero se sabe poco sobre la importancia de los componentes de señalización de

etileno en el control de este proceso de desarrollo. Un cribado de la plataforma de mutantes EMS para triple respuesta a etileno en plántulas etioladas permitió identificar cuatro mutantes insensibles a etileno dominantes o semidominantes: *ein1* (*etr1b*), *ein2* (*etr1a*), *ein3* (*etr2b*) y *EIN4* (*etr1a-1*). Las cuatro mutaciones alteraron los mecanismos de determinación del sexo. Las mutaciones *etr1a* y *etr2b* promovieron la conversión de flores femeninas en flores bisexuales y hermafroditas, y monoecia en andromonoecia, retrasando también la transición floral femenina y reduciendo el número de flores pistiladas por planta. Las mutaciones *etr1a-1* y *etr1b*, por otro lado, fueron capaces de bloquear por completo la transición floral femenina de *C. pepo*, haciendo que las plantas produjeran flores masculinas indefinidamente (androecia). Ya sea en homocigosis o heterocigosis, las cuatro mutaciones *etr* alteraron la tasa de crecimiento y la maduración de los pétalos y los carpelos de flores pistiladas, alargando el tiempo requerido para que las flores alcanzaran la antesis, y favoreciendo el desarrollo partenocárpico del fruto.

El análisis BSA mediante secuenciación de genoma completo (BSA-Seq) reveló que los genes subyacentes a las cuatro mutaciones *etr* codifican para receptores de etileno. Se descubrió que *etr1a*, *etr1a-1* y *etr1b* eran mutaciones de cambio de sentido en el dominio transmembrana de unión a etileno de los receptores CpETR1A y CpETR1B, mientras que *etr2b* era una mutación de cambio de sentido entre los dominios GAF e histidina quinasa del receptor de etileno CpETR2B. El nivel de insensibilidad al etileno en los mutantes simples y dobles homocigóticos y heterocigotos para las cuatro mutaciones *etr* estaba determinado por la fuerza de cada alelo mutante, pero también por el número de alelos WT y *etr* en la planta, lo que indica que los tres genes receptores de etileno no tienen una función totalmente redundante, sino que cooperan en el control de la respuesta de etileno. El nivel de insensibilidad a etileno determinará el fenotipo sexual de la planta, que va desde monoecia en plantas WT sensibles a etileno, pasando por andromonoecia en las plantas parcialmente insensibles a etileno, hasta androecia en las plantas más insensibles a etileno. Se observó que el bloqueo del programa de desarrollo femenino en los mutantes *etr1b* y *etr1a-1* estaba asociado con una falta de activación transcripcional de *CpACS11, CpACO2* y *CpACS2/7*, tres genes de biosíntesis de etileno necesarios para activar el desarrollo floral femenino, pero no cambiaba la expresión de *CpWIP1*, un factor de transcripción que parece ser necesario para el desarrollo de flores masculinas en otras especies de Cucurbitáceas. Se propone un modelo que integra la biosíntesis de etileno y los genes receptores de etileno en la red genética que regula la determinación del sexo en *C. pepo*.

Palabras clave: Etil-metanosulfonato (EMS), plataforma mutante, WGS, cribado masivo de mutantes, etileno, determinación sexual, cuajado de fruto, monoecia-andromonoecia-androecia, receptores de etileno.

Summary

The *Cucurbita pepo* genome comprises 263 Mb and 34,240 gene models organized in 20 different chromosomes. We generated an EMS mutant platform consisting of 3,751 independent M_2 families so as to improve our understanding of gene function. The quality of the collection has been evaluated based on phenotyping and whole-genome resequencing (WGS) results. The phenotypic evaluation of the whole platform at the seedling stage has demonstrated that the variation rate for easily-observable traits is more than 10%. The families with albino or chlorotic seedlings exceeded 3% of the platform, similar or higher to that found in other EMS collections of cucurbit crops. When four adult plants from each of 300 independent families were evaluated, more than 28% of apparent mutations were found to be associated with vegetative and reproductive traits, including: plant vigor; leaf size and shape; sex expression and sex determination; and fruit set and development. Two pools of genomic DNA, derived from 20 plants taken from each of two mutant families, were subjected to WGS by using NGS methodology so as to estimate the density, spectrum, distribution and impact of EMS-induced mutation. The number of EMS mutations in the genome of the L1 family was 1,704 and in the L2 one 859, which represents a density of 11.8 and 6 mutations per Mb, respectively. As expected, the predominant EMS-induced mutations were C>T and G>A transitions (80.3% in L1, and 61% L2), that were found to be randomly distributed along the 20 chromosomes of *C. pepo*. The mutations mostly affected intergenic regions, but 7.9% of the identified EMS mutations in L1 and 6% of those in L2, were located in the exome. 0.4% (L1) and 0.2% (L2) of the identified mutations had a moderate or high putative impact on gene functions. These results provide information regarding the potential usefulness of the obtained mutant platform in the search for novel alleles. This concerns both functional genomics and *Cucurbita* breeding, by using direct- or reverse-genetic approaches.

Ethylene is the key regulator of sex determination in the monoecious species of the family *Cucurbitaceae,* this hormone determines which individual floral meristem will develop into a female or a male flower. This is achieved by the arrest of stamen or carpel growth in early stages of flower development, and also during the transition between male and female developmental phases. Many of the regulatory genes discovered to date code for ethylene biosynthesis enzymes, but little is known of the importance of ethylene signaling components in the control of this developmental process. A high-throughput screening of the EMS platform for triple ethylene response in etiolated seedlings allowed the

identification of four semi-dominant or dominant ethylene-insensitive mutants: *ein1* (*etr1b*), *ein2* (*etr1a*), *ein3* (*etr2b*) and *EIN4* (*etr1a-1*). These four mutations modified sex determination mechanisms. The mutations *etr1a* and *etr2b* promoted the conversion of female into bisexual or hermaphrodite ones, and monoecy into andromonoecy. They also delayed the transition to female flowering and reduced the number of pistillate flowers per plant. The mutations *etr1a-1* and *etr1b*, on the other hand, were able to block female flowering transition of *C. pepo*; this induces plants to produce male flowers indefinitely (androecy). Whether in homozygous or heterozygous conditions, the four *etr* mutations disrupted the growth rate and maturation of petals and carpels of pistillate flowers, thereby lengthening the time required for flowers to reach anthesis and encouraging the growth rate of ovaries and the parthenocarpic development of fruits.

Bulked-segregant analysis coupled to whole-genome sequencing (BSA-Seq) revealed that the genes affected by the four *etr* mutations encode for ethylene receptor proteins. The *etr1a*, *etr1a-1* and *etr1b* were found to be missense mutations that resulted in an amino acid substitution in the ethylene-binding transmembrane domain of the ethylene receptors CpETR1A and CpETR1B, while *etr2b* was a missense mutation that affected to the coiled-coil domain between the GAF and histidine-kinase of the receptor CpETR2B. The level of ethylene insensitivity in homozygous and heterozygous single and double mutants was determined both by the strength of each mutant allele but also by the number of WT and *etr* alleles in the plant, indicating that the three ethylene receptor genes do not have a redundant function but cooperate in the control of ethylene response. The level of ethylene response in the plant determines its final sex phenotype: monoecy in ethylene-sensitive WT plants; andromonoecy in the partial ethylene-insensitive ones, and androecy in those which are the most highly ethylene insensitive. The blocking of female flowering transition by *etr1a-1* and *etr1b* -the two mutants which have the strongest effect- was found to be associated with a lack of transcriptional activation of the three ethylene biosynthesis genes *CpACS11*, *CpACO2* and *CpACS2/7*, which are required for female flower development. However, there was no change in the expression of *CpWIP1*, a transcription factor gene that seems to be required for male flower development in other cucurbit species. A model is proposed here, integrating the ethylene biosynthesis and receptor genes into the genetic network that regulates sex determination in *C. pepo*.

Keywords: Ethyl methanesulfonate (EMS), mutant platform, WGS, high-throughput screening, ethylene, sex determination, fruit set, monoecy/andromonoecy/androecy, ethylene receptors.

Abbreviation list

A

AI: Andromonoecy Index

ACC: 1-aminocyclopropane-1-carboxylate acid

ACS: ACC SYNTHASE

ACO: ACC OXIDASE

AF: Allele Frequency

AVG: Aminoethoxyvinylglycine

B

Bp: base pair

BLAST: Blast Alignment Search Tool

C

ºC: Degrees Celsius

CDPK: Calcium Dependent Protein Kinase

cDNA: complementary DNA

CTR1: Constitutive Triple Response 1

CuGenDB: Cucurbit Genomics Database

D

DNA: Deoxyribonucleic acid

DNAse: Deoxyribonuclease

DP: Read Depth

DPA: Days Post Anthesis

E

EBF: EIN3 Binding F-BOX

EDF: Ethylene-Response DNA-binding Factors

EF-1A: *Elongation factor 1-Alpha*

ein1: ethylene insensitive 1

ein2: Ethylene Insensitive 2

ein3: Ethylene Insensitive 3

EIN4: Ethylene Insensitive 4

EMS: Ethyl methanesulfonate

ERF: Ethylene Response Factor

ERS1/2: Ethylene Response 1/2

ETP1/2: Targeting Protein 1/2

ETR1/2: Ethylene Receptor 1/2

G

GA: Gibberellin

GATK: Genome Analysis Toolkit

G: grams

GQ: Genotype Quality

L

L1, L2: Mutant family 1 and mutant family 2

M

MAPK: Mitogen-Activated Protein Kinase

Mya: Million years ago

N

NCBI: National Center for Biotechnology Information

NGS: Next Generation Sequencing

No.: Number

P

PCR: Polymerase Chain Reaction

Q

q-PCR: quantitative PCR

R

RT-PCR: Real Time PCR

RNA: Ribonucleic Acid

mRNA: messenger RNA

S

SAM: S-adenosyl methionine

Subsp.: subspecies

STS: Silver Thiosulfate

SNP: Single Nucleotide Polymorphism

SNV: Single Nucleotide Variant

T

TILLING: Targeting-Induced Local Lesions in Genomes

U

UAL: University of Almería

UTR: Untranslated region

V

VCF: Variant Call Format

W

WGS: Whole-genome Sequencing

Table of contents

1. Introduction

1.1. *Cucurbita pepo*: origin and taxonomy

Cucurbita pepo L. is a dicotyledonous species with a chromosome number of 2n=2x=40 (Whitaker and Davis, 1962). This species belongs to the gourd family, *Cucurbitaceae*, which includes important vegetable crops cultivated worldwide such as cucumber (*Cucumis sativus* L.), melon (*Cucumis melo* L.), watermelon (*Citrullus lanatus* (Thunb.) Matsum & Nakai) and pumpkins, gourds and squash (*Cucurbita* L. *spp*.). *Cucurbita* was originated in the Americas and presents the highest diversity within the *Cucurbitaceae* family (Paris, 1986, 2000, 2016; Paris et al., 2015). This genus consists in 13 species, of which at least five are domesticated (*Cucurbita pepo*, *Cucurbita moschata* Duchesne, *Cucurbita maxima* Duchesne, *Cucurbita argyrosperma* Huber and *Cucurbita ficifolia* Bouché) (Bisognin, 2002). *Cucurbita pepo* is the most economically important species within the genus, is comprised of two subspecies, each of which includes four edible-fruited cultivar-groups or morphotypes; subsp. *pepo* includes Pumpkin, Vegetable Marrow, Cocozelle, and Zucchini and subsp. *ovifera* includes Acorn, Scallop, Crookneck, and Straightneck (Ferriol et al., 2003; Paris, 1986, 2000). Whole-genome resequencing of seven of these eight *C. pepo* morphotypes has been performed (Xanthopoulou et al., 2019), and has provided an improved understanding of the underlying genomic regions involved in the independent evolution and domestication of these two subspecies (Figure 1).

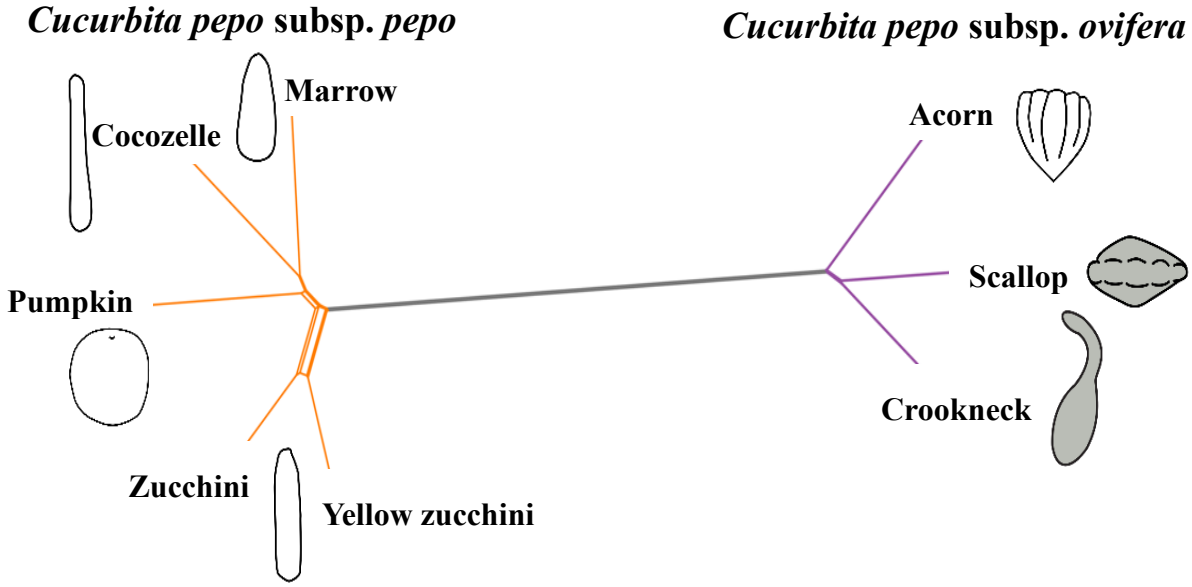


Figure 1. Phylogenetic network showing the relationship among the accessions described in *C.pepo* and schematic representation of their fruit shapes. Adapted from Xanthopoulou et al., 2019 and Paris, 1986.

Two independent domestication events have occurred in *Cucurbita pepo*, one (subsp. *pepo*) in Mexico and the other (subsp. *ovifera*) in the eastern United States (Paris, 1986). The earliest indications of the existence of Zucchini, the youngest of the *C. pepo* morphotypes, are from the latter half of the 19th century in northern Italy (Lust and Paris, 2016). This morphotype is now by far the economically most important of *C. pepo*.

1.2. Genomic resources in *Cucurbita pepo*

The increasing demand for fresh, high-quality food products has led to the development of new tools to generate new knowledge and obtain more sustainable fruits and vegetable varieties. Despite the importance of the genus *Cucurbita*, not many genetic and genomic tools were available until recently, preventing a better understanding of these species. However, in recent years, the rise of new sequencing technologies is helping to increase genetic knowledge so as to come to a clearer understanding of the biology of these species, as well as improving breeding programs of the *Cucurbita* crops. The main aims of zucchini breeding programs concern yield, together with fruit quality and resistance to disease.

Previous genetic studies of the *Cucurbita* genus indicated a possible whole-genome duplication (Esteras et al., 2012; Weeden, 1984). Recently, the *novo*-assembly of three *Cucurbita* genomes (*C. pepo, C. maxima* and *C. moschata*) has confirmed an allotetraploid origin of *Cucurbita* genus from two progenitors that diverged approximately 30 million years ago. Genomic studies showed that after the allotetraploidization event a low level of interchromosomal exchanges took place, suggesting high karyotype stability (Sun et al., 2017). However, differences in the expression patterns between the two subgenomes have been detected (Montero-Pau et al., 2017; Sun et al., 2017). This whole-genome duplication has not been observed in other species of the *Cucurbitaceae* family, such as cucumber (Huang et al., 2009), melon (Garcia-Mas et al., 2012) and watermelon (Guo et al., 2013). The *novo*-assembly and annotation of the *Cucurbita pepo* genome shows that it consists of 263 Mb and 27,868 gene models organized into 20 chromosomes. The complete annotated genome sequence of *C. pepo* is available, as well as the genome sequence of other cucurbit species, in the recently developed: Cucurbit Genomic Database. In addition, a total of 828 accessions of *C. pepo*, 294 of *C. moschata* and 355 of *C. maxima* have been genotyped using GBS technology. Single nucleotide polymorphisms (SNPs) have been discovered in each accession available, which permit to perform genotyping studies such as Genome Wide Association (GWAS). This tool will allow us to identify genomic regions involved in different plant processes (http://cucurbitgenomics.org/) (Zheng et al., 2019).

1.3. Mutant platforms and forward genetic

Chemically induced plant mutant populations have become an important source of variability for both functional genomic analyses and for plant breeding programs. These platforms can be generated using physical or chemical mutation. One of the most common mutant agents used is Ethyl methanesulfonate (EMS), a chemical mutagen that induces single randomly distributed nucleotide changes in DNA (Greene et al., 2003; Sega, 1984). Once established, the mutant platforms can be used for direct phenotyping and for DNA screenings, allowing the detection of single point mutations in a number of specific genes. 'Targeted Induced local Lesions IN Genomes' (TILLING) consists in the identification of mutations in genes of interest in populations treated with mutagenic agents (McCallum et al., 2000). The aleatory effect of this chemical mutagen is an advantage due to the variety of mutations that can be generated (synonymous, non-synonymous, splice acceptor or donor mutations, among others). However, it should be considered that this can also be inconvenient, since sometimes, in order to find the desired mutation, large mutant populations are necessary (Alonso and Ecker, 2006). TILLING was first used in Arabidopsis to detect allelic variants of a specific gene (Greene et al., 2003; McCallum et al., 2000). Later, this technology was extended in order to study a wide variety of organism, including animals. This tool was initially developed for functional genomics, but is also now being applied in plant breeding programs. TILLING technology goes beyond functional genomics and can be helpful in crop breeding programs (Slade and Knauf, 2005).

The proliferation of mutant platforms has occurred mainly in cultivated species of high agronomic interest, such as legumes, cereals and species of the *Solanaceae*, *Brassicaceae* and *Cucurbitaceae* families. Some examples of high-quality mutant collections developed include rice (Till et al., 2007), maize (Till et al., 2004), *Brassica napus* (Wang et al., 2008), tomato (Minoia et al., 2010), melon (Dahmani-Mardas et al., 2010; González et al., 2011) and cucumber (Boualem et al., 2014; Fraenkel et al., 2014), among others. In the case of zucchini, the first mutant platform was developed for assessing different target genes by using TILLING (Vicente-Dólera et al., 2014).

New useful mutants are being available thanks to high-throughput screenings based on direct plant phenotyping of large mutant collections of different crop species. The difficulty of direct screenings lies in the large number of plants to be evaluated for each of the traits under study. Massive screenings -based on phenotyping early seedling traits- have been developed for the detection of agronomically interesting mutants, including those with tolerance to

biotic and abiotic stresses, and also insensitive mutants to certain plant hormones or chemical treatments with specific herbicides or pesticides. Therefore, forward genetics is a powerful tool for revealing the genes that control developmental traits, as well as for revealing how these genes are connected in genetic networks (Mascher et al., 2014; Schneeberger, 2014).

1.4. Ethylene biosynthesis, perception and signaling pathways

Ethylene is a gaseous phytohormone that regulates multiple aspects of plant growth and development. Ethylene exerts its action at very low concentrations and occurs in all tissues of the plant, modulating its production at different plant stages (Argueso et al., 2007). In different plants species, ethylene controls, among other processes: seed germination, dormancy rupture, plant growth, root nodulation, cellular respiration, sexual expression, fruit maturation and abscission, and flowers and fruits senescence (Abeles et al., 2012). Due to its importance in plants, the molecular control of the biosynthesis, perception and signaling pathways has been subject to intensive study.

1.4.1. Ethylene biosynthesis

The first step in the ethylene biosynthesis pathway is the formation of S-adenosyl methionine (SAM) by the enzyme SAdoMet synthetase (Yang and Hoffman, 1984). This component is used as substrate for many biochemical biosynthesis pathways, including polyamines and ethylene. The most critical step in ethylene biosynthesis is the conversion of SAM into 1-aminocyclopropane-1-carboxylate (ACC), catalyzed by ACC Synthase (ACS). In the final step, ACC is oxidized by ACC oxidase to form ethylene, producing CO2 and cyanide also (Adams and Yang, 1979; Johnson and Ecker, 1998; Mirica and Klinman, 2008).

ACC synthase and ACC oxidase:

The ACC synthase enzymes are encoded by a multigene family in many plant species. In the *Cucurbitaceae* species, several different genes have been cloned which are highly homologous with the isoforms described in Arabidopsis and tomato (Boualem et al., 2008, 2009, 2015; Nakajima et al., 1990; Trebitsh et al., 1997). In cucumber, six *ACS* genes have been identified: *CsACS1*, *CsACS1G*, *CsACS2*, *CsACS3*, *CsACS4*, and *CsACS11* (Boualem et al., 2015, 2016; Kamachi et al., 1997; Shiomi et al., 1998; Trebitsh et al., 1997). *ACS* genes were first characterized in zucchini by Sato and Theologis (1989) and in tomato by (Van der Straeten et al., 1990). The recent discovery of a complete duplication in the genus *Cucurbita* (Montero-Pau et al., 2017; Sun et al., 2017) has allowed the identification of at

least two paralogous for most of *ACS* genes, although it has also been revealed that only one of the copies has maintained its function (Huang et al., 1991; Martínez et al., 2014).

ACC oxidases (ACO) are encoded by a multigene family, but this is made up of a lower number of members than the *ACS* multigene family. Five *ACO* genes have been found in Arabidopsis, and 6 genes in tomato (Lasserre et al., 1996; Ruduś et al., 2013). Four *ACC oxidase* genes have been isolated in cucumber: *CsACO1*, *CsACO2*, *CsACO3* (Kahana et al., 1999) and *CsACO4* (Chen, 2012). Orthologous genes have been identified, containing each of them a paralogous in the duplicated genome of *C. pepo*. As observed for *ACS* genes, only one of each paralogous *ACO* genes maintains its expression and function.

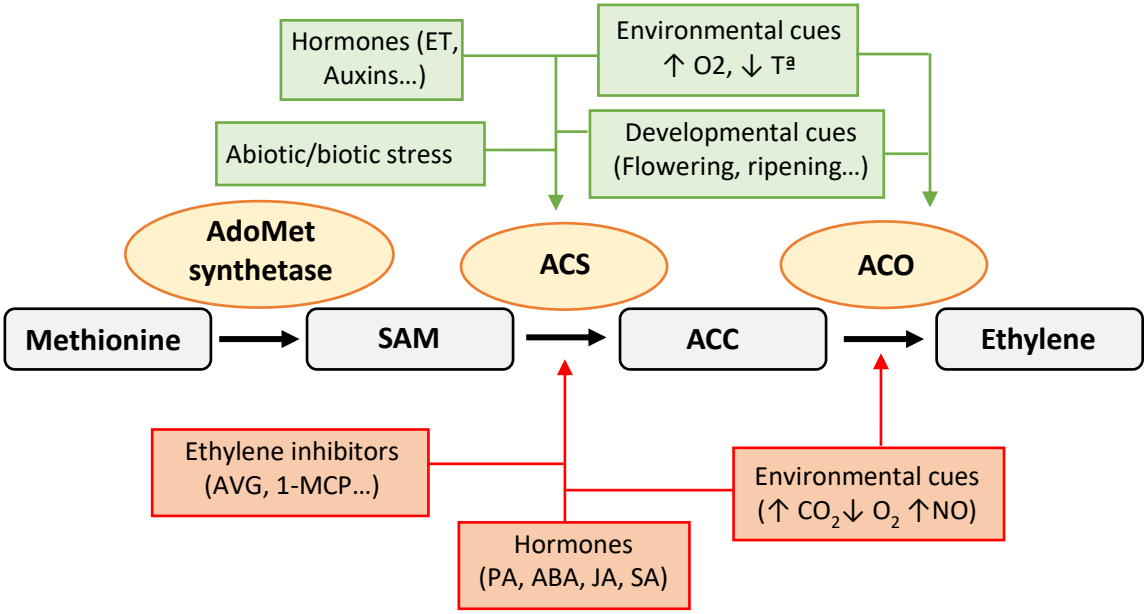


Figure 2. Regulation of the ethylene biosynthesis pathway. The enzymes that mediate reactions are ACC synthase (ACS) and ACC oxidase (ACO). The promoting and blocking agents for the two enzymes are indicated in green and red, respectively. ET, ethylene; AUX, auxins; ABA, abscisic acid; BR, brassinosteroids; Tª, temperature; O_2, oxygen. Adapted from Argueso 2007; Martínez 2013; Megías 2016.

The members of the ACS an ACO gene families are differentially regulated at transcriptional level and may be induced by a variety of internal signals, including ethylene itself (Bleecker and Kende, 2000; Li et al., 2012; Rottmann et al., 1991; Salman-Minkov et al., 2008). Thus, a diverse group of factors enhances ethylene production by modulating the transcription or the activity of ACS and ACO, including abiotic and biotic stresses, as well as hormonal, genetic and environmental stimuli (Argueso et al., 2007). The application of hormones such

as auxins and brassinosteroids can affect ethylene biosynthesis (Abeles et al., 1992; Vogel et al., 1998; Woeste et al., 1999), and ethylene precursors can also modify the expression of some *ACS* genes (Argueso et al., 2007; Barry et al., 2000; Li et al., 2012) (Figure 2).

1.4.2. Ethylene perception

Ethylene perception and signaling genes have been identified by the analysis of mutants with altered ethylene response in Arabidopsis (Stepanova and Alonso, 2009). When seedlings are grown in darkness and in the presence of ethylene, they display a characteristic triple response: decrease in the length of hypocotyl and roots, an increased hypocotyl thickness and greater curvature of the apical hook. The identification and characterization of mutants that were altered in ethylene triple response allowed the molecular bases of ethylene-regulated processes to be discerned (Bleecker et al., 1998).

The response of seedlings to ethylene has been successfully used to screen different EMS mutant collections in Arabidopsis (Guzmán and Ecker, 1990), and to identify many mutants affected in the ethylene pathway (Roman et al., 1995). The analysis and characterization of those mutants, such as ethylene receptor (*etr*) and ethylene insensitive (*ein*) ones, has allowed the cloning of five ethylene receptor genes: *AtETR1*, *AtETR2*, *AtERS1*, *AtERS2* and *AtEIN4* (Hall et al., 2000; Sakai et al., 1998), which, in their functional state, form dimers and higher-molecular weight oligomers at the endoplasmic reticulum (ER) membrane (Chen et al., 2010).

Mutations in ethylene receptor genes fall into two main categories: (i) dominant gain-of-function mutations, conferring ethylene insensitivity, and (ii) recessive loss-of-function mutations that have little effect as single mutations, but show a constitutive ethylene response in combination with other mutations, for example, in double, triple, and quadruple mutants of Arabidopsis. In dominant mutants, the ethylene-insensitive phenotypes are caused by single amino acid substitutions in the transmembrane ethylene-binding domain of any of the five ethylene receptors described in this species (Bleecker et al., 1988; Chang et al., 1993; Guzmán and Ecker, 1990; Hua et al., 1995; Hua and Meyerowitz, 1998; Wang et al., 2006). These mutational studies have contributed to understanding relation structure-function, where different classes of loss-of-function mutants showed different levels of ethylene binding, signal transmission, and intrinsic activity depending on the change position in the protein (Schott-Verdugo et al., 2019; Wang et al., 2006). Receptors are negative regulators of the ethylene response, and their expression level seems to be inversely

correlated with tissue ethylene-sensitivity (Klee, 2004). Other important characteristics of ethylene perception genes have been demonstrated through the analysis of double (*ers1etr1*) and triple (*etr2ers2ein4* and *etr1etr2ein4*) mutants. Ethylene receptors are functionally redundant (Hua and Meyerowitz, 1998), and single and double mutants are still able to respond to ethylene.

Based on their protein structure, two subfamilies of ethylene receptors have been described in Arabidopsis: ETR1/ERS1 (subfamily I) and ETR2/ERS2/EIN4 (subfamily II) (Figure 3). The N-terminal ethylene binding domain of subfamily I contains three transmembrane subdomains, while those of subfamily II presents four. The cytoplasmic portion of the receptors is made up of a GAF-related domain of unknown function, a histidine protein kinase domain, and a receiver domain, which is lacking in ERS1 and ERS2 (Kieber et al., 1993). In vitro studies indicate that subfamily I receptors have Histidine kinase activity, while subfamily II members and ERS1 have Ser/Thr kinase activity (Gamble et al., 1998; Moussatche and Klee, 2004). It has been suggested that the members of subfamily I initiate the ethylene perception process, whereas members of subfamily II intensify the signal, interacting with subfamily I members (Cancel and Larsen, 2002; Chang and Stadler, 2001; Gamble et al., 2002).

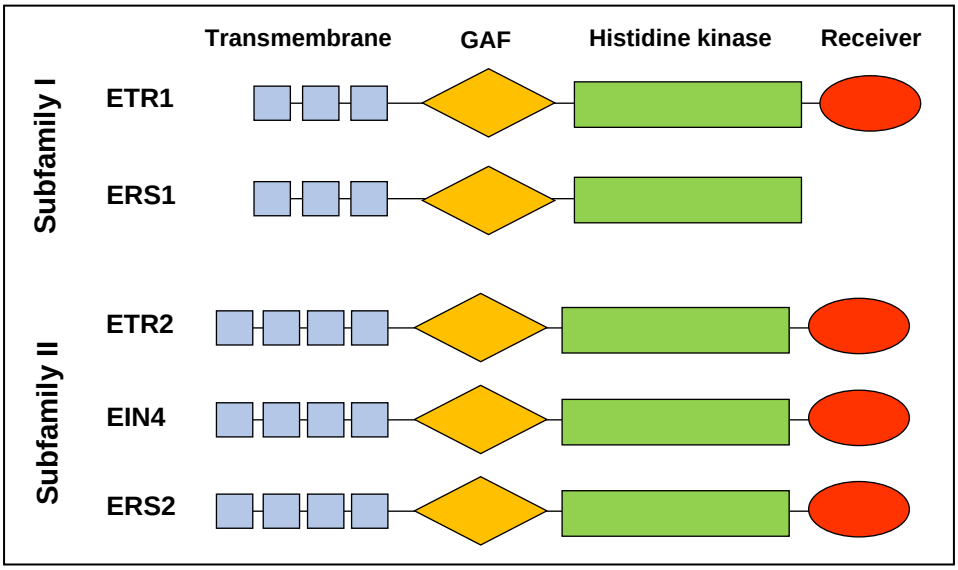


Figure 3. Ethylene receptors in Arabidopsis. All members of the receptor family have a similar overall modular structure composed of an N-terminal transmembrane sensor domain (TMD) (blue), a GAF domain in the middle portion (yellow), and a catalytic transmitter domain at the C-terminus (green). In receptors ETR1, ETR2 and EIN4, this basic structure is complemented by a C-terminal receiver domain (red).

Recently, the first structural model of the transmembrane sensor domain of the plant ethylene receptor ETR1 in Arabidopsis has been published. This research highlights the important role of mutational studies to reach a better understanding of the activity of the different parts of the receptor, especially explaining how ethylene can bind proximal to the copper ions in the receptor (Schott-Verdugo et al., 2019).

1.4.3. Ethylene signal transduction pathway

The ethylene transduction pathway is initiated when the hormone is perceived by a family of histidine-kinase receptors that are negatively regulated by ethylene. The interaction between ethylene and the binding sites of the receptors is mediated by Cu^{2+} (Rodriguez et al., 1999), which is transferred to the receptors by the protein RESPONSIVE TO ANTAGONIST 1 (RAN1) (Hirayama et al., 1999; Woeste and Kieber, 2000). The Ethylene Receptors 1 and 2 (ETR1 and ETR2), ethylene response (ERS1 and ERS2) and Ethylene Insensitive 4 (EIN4), together with the Raf-like kinase CONSTITUTIVE TRIPLE RESPONSE (CTR1), function as negative regulators of the ethylene pathway, and in absence of ethylene, they repress downstream positive components of the signal transduction pathway. Active CTR1 phosphorylates the C-terminal domain of Ethylene Insensitive 2 (EIN2), preventing its nuclear localization and making it a target for 26S proteasomal degradation by F-box proteins EIN2 TARGETING PROTEIN 1/2 (ETP1/2). The union of ethylene locks the receptors in an inactive conformation, relieving the constitutive repression of ethylene response (Stepanova and Alonso, 2009; Wen et al., 2012).

The stability of ETHYLENE INSENSITIVE3 (EIN3) is regulated via phosphorylation of a specific threonine residue by an MKK9–MPK3/6 cascade. In the presence of ethylene, the F-box proteins EIN3 BINDING F-BOX (EBF1 and EBF2) fail to degrade EIN3 and EIL1. As a result, these transcription factors are accumulated in the nucleus, activating primary target genes, such as ETHYLENE RESPONSE FACTORS1 (ERF1), ETHYLENE-RESPONSE DNA BINDING FACTORS (EDF) and EBF2, and triggering the ethylene response.

A negative regulator of ethylene response, CTR1, acts downstream of the ethylene receptors; this is a protein homologue to the family RAF LIKE MITOGEN ACTIVATED PROTEIN KINASE KINASE KINASES (MAPKKK) (Clark et al., 1998; Huang et al., 2003; Kieber et al., 1993). Such proteins are involved in various external and developmental signal transduction pathways (Kyriakis et al., 1992; Pelech and Sanghera, 1992). The constitutive

triple response phenotype of the Arabidopsis recessive mutant *ctr1* indicates that *CTR1* is a repressor of ethylene response (Kieber et al., 1993). CTR1 acts downstream of receptors and upstream of EIN2 (Roman et al., 1995). Consequently, CTR1 activity is likely regulated by its association/dissociation with the receptors ETR1, ETR2 and ERS1, being the interaction with ETR1 the strongest (Cancel and Larsen, 2002; Clark et al., 1998; Gao et al., 2003). Although CTR1 gene is unique in Arabidopsis (Kieber et al., 1993), a multigene family of four members has been identified in tomato, and the products of this family appear to be involved in a number of processes in different tissues (Alexander and Grierson, 2002; Gao et al., 2003; Leclercq et al., 2002; Lin et al., 1998, 2008). In the *Cucurbitaceae* family, five *CTR*-like genes have been cloned and characterized in *C. melo*, *C. sativus* and *C. pepo* (Manzano et al., 2010b).

One of the last steps in the ethylene signal transduction pathway is covered by the *ETHYLENE INSENSITIVE 2* (*EIN2*) gene of Arabidopsis. The mutation *ein2* inhibits triple ethylene response of etiolated seedlings (McGrath and Ecker, 1998), indicating that EIN2 is a positive regulator of ethylene response. It is likely that EIN2 acts in several signal transduction pathways in response to various hormones (Alonso et al., 1999). The ETR1-CTR1 complex regulates EIN2 activity by a cascade of MAPK phosphorylation that terminates in EIN2 (Ouaked et al., 2003). The posttranscriptional regulation of EIN2 seems to be mediated by the F-box proteins EIN2 TARGETING PROTEIN1/2 (ETP1/2) (Stepanova and Alonso, 2009). In absence of ethylene, EIN2 is phosphorylated by the ETR1-CTR1 complex and targeted for 26S proteasomal degradation by F-box proteins ETP1/2 (Stepanova and Alonso, 2009). The EIN2 signaling converges in EIN3 and EIN3-like transcription factors. These nuclear proteins are transcription factors that activate the transcription of the genes involved in ethylene response (Chao et al., 1997; Stepanova and Alonso, 2009). Consequently, *ein3* recessive mutant shows reduced response to ethylene, lack of the triple response, reduced expression of ethylene dependent genes and delayed leaf senescence (Solano et al., 1998).

In response to ethylene, EIN3 is an essential regulator in the pathway of signal transduction. EIN3 and EIN3-like can bind specific sequences in the DNA (Solano et al., 1998), altering the expression of direct targets such as ETHYLENE RESPONSE FACTOR1 (ERF1) (Solano et al., 1998) and EIN3 BINDING F-BOX (EBF2) (Konishi and Yanagisawa, 2008). As occurs with EIN2, transcription factors EIN3/EIL1 are also targeted for 26S proteasomal degradation, a post-transcriptional mechanism dependent on the F-box proteins EBF1 and

EBF2 (Guo and Ecker, 2004; Lee et al., 2006; Potuschak et al., 2003). EIN3 is also regulated by the cytoplasmic exonuclease EIN5/XRN4 (Kastenmayer and Green, 2000). The characterization of a classical ethylene insensitive mutant, *ein5*, indicated that the *EBF2* mRNA levels were upregulated (Olmedo et al., 2006; Potuschak et al., 2006), suggesting that EIN5 promotes mRNA decay of the F-box genes EBF1/2, leading to higher levels of EIN3.

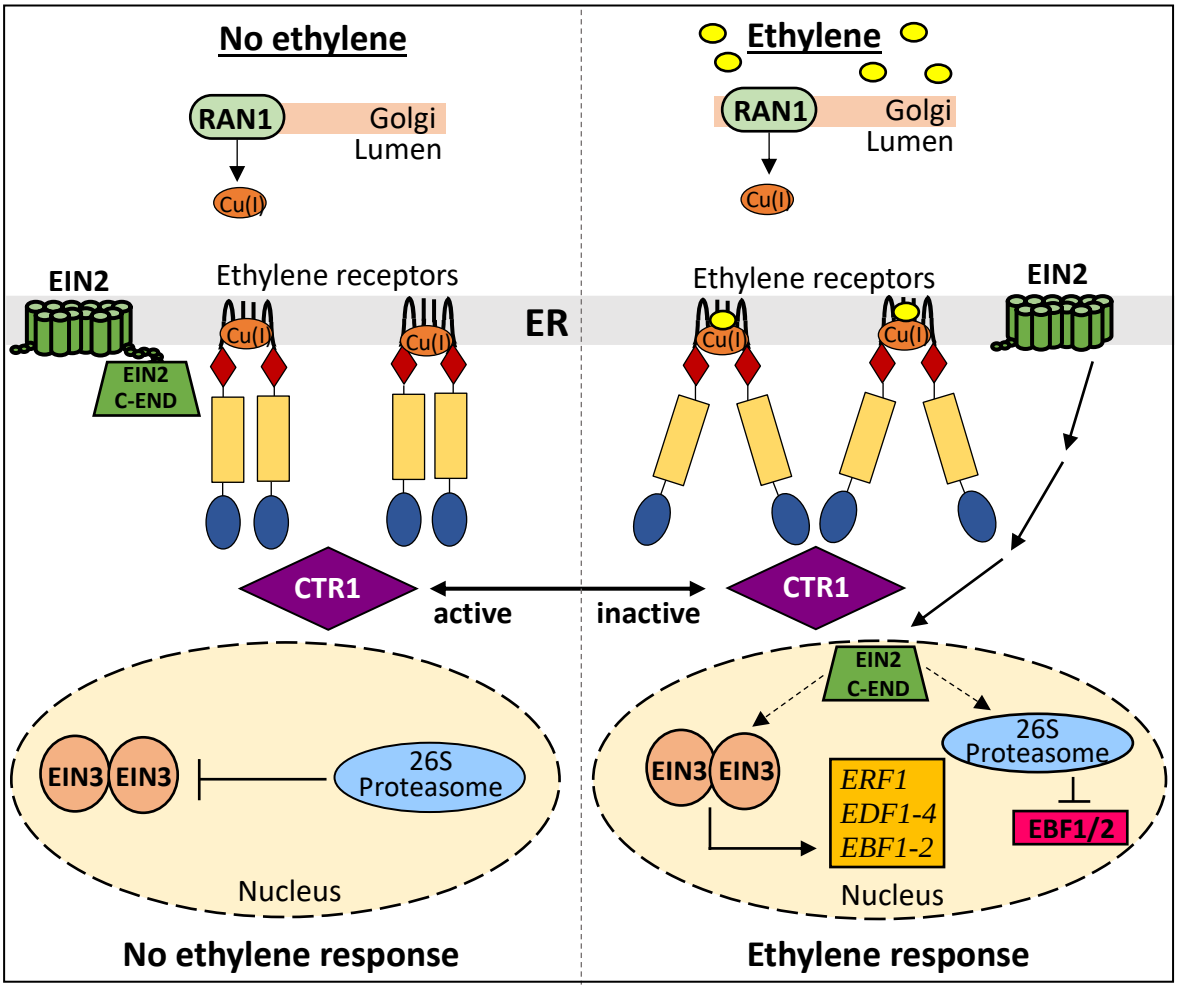


Figure 4. Ethylene signal transduction pathway in Arabidopsis. Ethylene is perceived by 5 receptors located in the endoplasmic reticulum (ER) membrane. RAN1, a copper transporter, is involved in the delivery of copper ions to receptors to facilitate ethylene binding. Receptors act as negative regulators of ethylene response. In the absence of ethylene, receptors bind to the Raf-like kinase CTR1, which regulates negatively the downstream ethylene response pathway, by phosphorylating the C-terminal domain of EIN2 and through a MAP-kinase that terminates in the EIN3/EIL family of transcription factors. The phosphorylation of EIN2 and EIN3/EIL represses their activity by promoting their 26S proteasomal degradation. When ethylene binds to receptors, the receptor complexes disassociate and CTR1 is inactivated and degraded by the proteasome. The inactivation of the phosphorylation cascade therefore results in the accumulation of EIN3 and EIN3-like (EILs) in the nucleus, where EIN3 binds to the promoter of ERF1 gene and activates its transcription in an ethylene-dependent manner. Transcription factors ERF1, EDF1-4 and EBF1-2 act as the starting point of a new transcriptional cascade that results in the activation and repression of a number of genes involved in ethylene response.

1.5. Sex expression and sex determination in Cucurbits

Most flowering plants are hermaphrodite, developing bisexual flowers with both reproductive organs, stamens and carpels. Only a small proportion of plant species have adopted monoecy or dioecy as reproductive strategy, producing female or male unisexual flowers in the same plant or in separate plants, respectively. The development of unisexual flowers is the result of a modification in the development pattern of a hermaphrodite flower, indicating that dioecious and monoecious plants have evolved from hermaphrodite ancestors (Sun et al., 2010). In cucurbits, other flowering morphotypes have been also described: andromonoecious (plants with male and hermaphrodite flowers), gynoecious (plants with female flowers only), androecious (plants with male flowers only), or trimonoecious (plants with male, female and hermaphrodite flowers) (Pannell, 2017; Perl-Treves, 2004) (Figure 5).

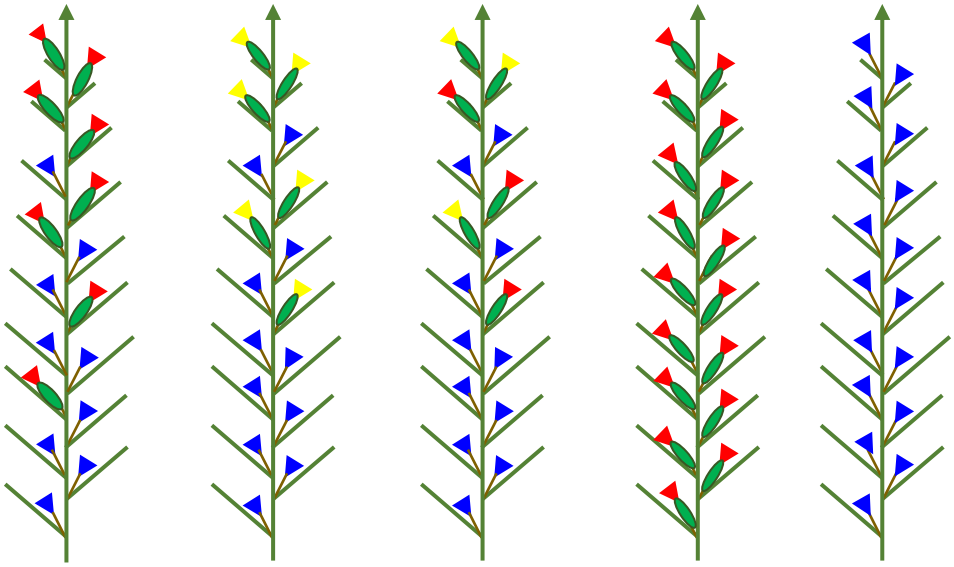


Figure 5. Sexual morphotypes observed in Cucurbits. Schematic representation of the distribution of male and female flowers in plants. Blue, male flower; red, female; green, bisexual; yellow, hermaphrodite.

1.5.1. Environmental and hormonal factors involved in sex expression and sex determination

Sex expression, that is the number and distribution of male and pistillate flowers per plant, is regulated by the interaction of environmental, genetic and hormonal factors (Diggle et al., 2011), ethylene being the main regulator. With regard to environmental factors, it has been reported that winter conditions with short days, low light intensity and low night

temperatures, promote female flower production, while summer conditions with longer days, and higher light intensity, are able to induce the production of male flowers (Peñaranda et al., 2007; Rudich et al., 1972b; Wien et al., 2004).

In addition to environmental conditions, hormones also mediate flower development in cucurbits. Hormonal effects have also been studied through external applications (Perl-Treves, 2004; Rudich et al., 1972a). The application of gibberellins (GAs) promotes stamen development in female flowers of monoecious lines in cucumber and melon (Galun, 1959; Peterson and Anhder, 1960; Shifriss and George, 1964), whereas inhibitors of GAs promote feminization in cucumber or zucchini (Yin and Quinn, 1995). Contrary to GAs, auxins increase the percentage of female flowers in cucumber monoecious lines (Galun et al., 1965), as well as in melon and zucchini (Galun, 1959; Rudich et al., 1969). The effects of these two hormones appear to be modified by endogenous ethylene, whose level is altered by the application of GAs and auxins (Ando et al., 2001; Byers et al., 1972a). Brassinosteroids can also increase the percentage of female flowers in cucumber and squash, although their action seems also to be modified by an increase in ethylene production (Manzano et al., 2011; Papadopoulou and Grumet, 2005).

Ethylene is the key hormone regulating sex expression in cucurbits. The exogenous application of ethylene causes feminization in cucumber, melon and zucchini (Manzano et al., 2010b; Mcmurray and Miller, 1968; Owens et al., 1980; Rudich et al., 1969), whereas the application of ethylene inhibitors such as Aminoethoxyvinylglycine (AVG) and silver thiosulfate (STS) increases the number of bisexual or male flowers in gynoecious and monoecious lines of cucumber (Byers et al., 1972b) and squash (Manzano et al., 2010b, 2014; Martínez et al., 2013). The level of endogenous ethylene also correlates with sex expression in the genus *Cucumis*. Thus, ethylene production in seedlings is 2-3 times higher in gynoecious than in monoecious lines of cucumber (Rudich et al., 1972b; Trebitsh et al., 1987; Yamasaki et al., 2001, 2003). Contrary to what is observed in other cucurbits, ethylene promotes male flower development in watermelon, favoring an increase in the percentage of males in the whole plant (Manzano et al., 2014).

Ethylene also regulates cucurbit sex determination, that is, the fate of each floral meristem in the plant. Ethylene determines whether the floral meristem will become a female of a male flower. The development of a female flower requires a higher level of ethylene than the development of a male flower (Manzano *et al.*, 2010, 2011, 2014; Zhang *et al.*, 2017). If the production of ethylene in the floral meristem is high, this arrests the development of stamen

primordia, and the meristem is developed into a female flower. When the production of ethylene is reduced, the carpel is arrested in their development, and the floral meristem is become into a male flower (Kater et al., 2001; Malepszy and Niemirowicz-Szczytt, 1991). Under treatments that reduce ethylene biosynthesis or perception, the monoecious plants are converted into partially or completely andromonoecious ones, demonstrating that ethylene participates in the sex identity of female flowers, and that an ethylene threshold is required to arrest stamen development in the female flower (Manzano *et al.*, 2011; Zhang *et al.*, 2017). For example, ethephon applications in androecious mutant plants can revert the sexual phenotype to monoecious by inducing female flower development in cucumber (Chen et al., 2016)

1.5.2. Genetic and molecular control of sex expression and sex determination in Cucurbits

The main stem of monoecious cucurbits is characterized by three phases of floral development. In the initial phase, male flowers are produced at lower nodes. In the second phase, male and female flowers are formed in the middle part of the plant, while female flowers prevail at the upper nodes during the third phase. Genetic studies in cucumber and melon have determined that these different sex morphotypes are controlled by four main loci, *F*, *M*, *A*, and *Gy* (Malepszy and Niemirowicz-Szczytt, 1991). These loci have been identified and characterized, and most of them corresponds to genes that encode for ethylene biosynthesis enzymes ACS and ACO (Table 1).

Locus F/f (Female): Homozygous plants containing the dominant *F* allele are female. This locus has been identified as an additional nearby copy of the *CsACS1* gene, encoding for 1-aminocyclopropane-1-carboxylic acid synthase (ACS1), and was designated as *CsACS1G* (Knopf and Trebitsh, 2006; Mibus and Tatlioglu, 2004; Trebitsh et al., 1997). Thus, the feminizing effect of *F* allele could be a dose effect which leads to an increase in ethylene production. This would explain the high endogenous level of ethylene in gynoecious lines (Knopf and Trebitsh, 2006; Shiber et al., 2008).

Locus M/m (Monoecious): The dominant *M* allele confers monoecy, while the recessive allele confers andromonoecy. The epistatic interaction of *F* and *M* loci determine that plants with the genotype *mmff* are andromonoecious and those with *mmF_* are hermaphrodite. Flowers of plants harboring *M_F_* are female, whereas *M_ff* plants are monoecious with mostly male flowers. The M/m locus has been cloned in cucumber and shown to be another

member of the *ACS* gene family called *CsACS2*. The m allele has a mutation (Gly33Cys) at a conserved site of the protein (Li et al., 2009b). In melon, watermelon and squash the orthologous genes have been also identified corresponding to *CmACS7*, *CitACS4* or *CpACS27A*, respectively (Boualem et al., 2009; Ji et al., 2016; Manzano et al., 2016; Martínez et al., 2014). *CsACS2* and *CmACS7* are female specific genes that are expressed when the pistil begins to develop (Saito et al., 2007), and the time of its expression is consistent with the action of ethylene during the induction of the first female flower in monoecious and gynoecious lines (Kamachi et al., 1997). In squash, *CpACS27A* is expressed at an early stage of pistillate flower development and is progressively downregulated until the stage of complete development is reached (Martínez et al., 2014).

Locus A/a (androecious): Homozygous plants for the *a* allele are characterized by the masculinization of the plants and is hypostatic to the *F* gene. Plants with the genotypes *mmffaa* and *M_ffaa* are completely androecious. As described in Boualem *et al.*, 2015, this gene corresponds to *CsACS11* in cucumber and *CmACS11* in melon. To the date, there are no other orthologues identified in watermelon or squash. Moreover, mutations in the ethylene biosynthesis gene *CsACO2* lead to androecy in cucumber, indicating that this gene is also involved in carpel development and probably collaborates with *ACS11* in this function (Chen et al., 2016).

Locus Gy/gy (gynoecious): The recessive allele *gy* is responsible for gynoecy. The gene encodes for *CsWIP1* in cucumber and *CmWIP1* in melon, a transcription factor that regulates carpel abortion during the development of male flowers, and the mutations in this gene leads to gynoecy. Plants lacking *WIP1* expression, or when *WIP1* is dysfunctional, produce only female flowers (Boualem et al., 2015; Chen et al., 2016; Hu et al., 2017; Martin et al., 2009). Recently, the orthologous *ClWIP1* gene has also been identified in watermelon, where it has the same function as in melon and cucumber (Zhang et al., 2019). *WIP1* expression is repressed by *ACS11*, allowing the coexistence of male and female flowers in the plant (Boualem et al., 2015; Chen et al., 2016).

There are evidences indicating that ethylene receptors also play a key role in sex determination, although most of the information we have to date derives from gene expression data in different sex morphotypes or from transgenic plants in heterologous systems. The expression of *CsETR1*, *CsETR2* and *CsERS* in the apical shoots of cucumber is known to be higher in gynoecious than in monoecious lines, and their expressions are upregulated in the transition to female flowering (Yamasaki et al., 2001). This indicates that

ethylene perception plays an active role in the activation of the female flowering program. Downregulation of *CsETR1* appears to be required for the arrest of stamen development during female flower development (Wang et al., 2010a). Moreover, when the Arabidopsis ethylene-insensitive allele *etr1-1* is over-expressed in transgenic *C. melo* plants, sex determination and sex expression are modified (Little et al., 2007; Switzenberg et al., 2015).

Table 1. List of genes involved in sex determination in different cucurbit species.

Gene	Protein	Function in sex determination	LOF mutations	References
ACS1G	Additional ethylene biosynthesis enzyme	Activation of the female flower development	-	(Knopf and Trebitsh, 2006; Mibus and Tatlioglu, 2004; Trebitsh et al., 1997)
ACS2/7	Ethylene biosynthesis enzyme	Arrest of stamen development during female flower development	Andromonoecy	(Boualem et al., 2009; Ji et al., 2016; Manzano et al., 2016; Martínez et al., 2014)
ACS11	Ethylene biosynthesis enzyme	Promotion of carpel development	Androecy	(Boualem et al., 2015)
WIP1	Transcription factor	Arrest of carpel development during male flower development	Ginoecy	(Boualem et al., 2015; Chen et al., 2016; Hu et al., 2017; Martin et al., 2009; Zhang et al., 2019)
ACO2	Ethylene biosynthesis enzyme	Promotion of carpel development	Androecy	(Chen et al., 2016)
ETR1, ETR2, ERS1	Ethylene receptor and signaling	Carpel development and stamen arrest	-	(Little et al., 2007; Sun et al., 2010; Switzenberg et al., 2015; Wang et al., 2010a)
ERF110	Ethylene response factor 110	Induces *ACS11* expression and feminization	-	(Tao et al., 2018)
ERF025	Ethylene response factor 025	May promote feminization tendency, upregulates *ACS* and *ACO* genes	-	(Wang et al., 2017)
ERF31	Ethylene response factor 31	Promote M gene expression and feminization	-	(Pan et al., 2018)

Recently, other ethylene signaling genes called ethylene responsive factors (*ERF*) have been identified, which modulate the sexual phenotype of the plants by activating ethylene response. Overexpression of *ERF025* increases ethylene production, upregulating *ACS* and *ACO* ethylene biosynthesis genes and consequently influencing sex determination (Wang et al., 2017). *ACS2* and *ACS11*, identified as sex determining genes, have been found to be directly regulated at a transcriptional level by *ERF110* (Tao et al., 2018). In addition, biochemical experiments using yeast one-hybrid assays have demonstrated that *ERF31* could bind the *M* promoter, thereby activating its expression. *CsERF31* responded to the ethylene signal derived from *CsACS1* and could mediate the positive feedback regulation of ethylene by activating *M* expression (Pan et al., 2018). Despite sex determination and sex expression in cucurbits species have been subjected to several studies, the complete molecular regulation of these processes is still unclear.

2. Objectives

Ethylene is the key regulator of sex determination in the monoecious species of the *Cucurbitaceae* family. This hormone regulates the number and distribution of female and male flowers in the plant, as well as the development of each floral meristem. Although it has been identified a number of *ACS* and *ACO* genes that play an essential role in this development process, the involvement of ethylene perception and signaling genes in sex determination is much less well known, especially in *Cucurbita pepo*.

The main objective of this work is to gain insight into the genetic control of sex determination in *C. pepo*, combining high-throughput screenings and next-generation sequencing technologies (NGS) so as to determine the role played by ethylene signaling genes in sex expression and sex determination. To this end we selected the following specific objectives:

1. Characterize a platform of 3,751 EMS mutant lines of *C. pepo* at genomic and phenomic levels, so as to: estimate the density, spectrum, distribution and impact of EMS-induced mutation; evaluate the phenotype of the whole platform at seedling stage; provide information regarding the potential use of the mutant platform in the search for novel undiscovered alleles, for use in both functional genomics and *Cucurbita* breeding.

2. Identify ethylene-insensitive mutants by using the triple response test to ethylene in etiolated seedlings, so as to study their inheritance patterns and alterations in their phenotypes during plant development.

3. Identify the causal mutations of the mutant phenotypes by using whole-genome sequencing (WGS), and to determine the molecular structure and function of the genes involved, and also the transcriptional regulation of other sex-determining genes.

3. Phenomic and genomic characterization of a mutant platform in *Cucurbita pepo*

3.1. Abstract

The *Cucurbita pepo* genome comprises 263 Mb and 34,240 gene models organized in 20 different chromosomes. To improve our understanding of gene function we have generated an EMS mutant platform, consisting of 3,751 independent M_2 families. The quality of the collection has been evaluated based on phenotyping and whole-genome re-sequencing (WGS) results. The phenotypic evaluation of the whole platform at seedling stage has demonstrated that the rate of variation for easily observable traits is more than 10%. The percentage of families with albino or chlorotic seedlings exceeded 3%, similar or higher to that found in other EMS collections of cucurbit crops. A rapid screening of the library for triple ethylene response in etiolated seedlings allowed the identification of four ethylene-insensitive mutants, that were found to be semi-dominant (*ein1*, *ein2* and *ein3*) or dominant (*EIN4*). By evaluating 4 adult plants from 300 independent families more than 28% of apparent mutations were found for vegetative and reproductive traits, including plant vigor, leaf size and shape, sex expression and sex determination, and fruit set and development. Two pools of genomic DNA derived from 20 plants of two mutant families were subjected to WGS by using NGS methodology, estimating the density, spectrum, distribution and impact of EMS induced mutation. The number of EMS mutations in the genomes of families L1 and L2 was 1,704 and 859, respectively, which represents a density of 11.8 and 6 mutations per Mb, respectively. As expected, the predominant EMS induced mutations were C>T and G>A transitions (80.3% in L1, and 61% L2), that were found to be randomly distributed along the 20 chromosomes of *C. pepo*. The mutations were mostly affecting intergenic regions, but 7.9% and 6% of the identified EMS mutations in L1 and L2, respectively, were located in the exome, and 0.4 and 0.2% had a moderate and high putative impact on gene functions. These results provide information regarding the potential use of the obtained mutant platform in the discovery of novel alleles for both functional genomics and *Cucurbita* breeding by using direct- or reverse-genetic approaches.

Keywords: *Cucurbita pepo*, EMS, mutant platform, WGS, high-throughput screening, ethylene.

3.2. Introduction

The genus *Cucurbita* comprises three important worldwide-cultivated crop species: *C. pepo, C. maxima* and *C. moschata*. The most important cultivated species of the genus is *C. pepo*, consisting of different morphotypes of summer and winter squashes and gourds with great

economic importance such as Zucchini, Pumpkin, Vegetable Marrow, Cocozelle, Scallop, Acorn, Straightneck and Crookneck. In the year 2014 the cultivated area of *C. pepo* reached nearly 2 million of hectares and a production of 25 million of tons (FAOSTAT).

The genome of *C. pepo* has been recently sequenced and annotated (cucurbitgenomics.org; Montero-Pau et al., 2017). It consists of 263 Mb, a scaffold N50 of 1.8 Mb, and 34,240 gene models, organized in 20 chromosomes that cover about 93% of the assembled sequence. A duplication event has been detected in the genome of *Cucurbita*, happening about 20 Mya ago, after the separation from other genus of the *Cucurbitaceae* family such as *Citrullus* (watermelon) and *Cucumis* (melon and cucumber) (Montero-Pau et al., 2017; Sun et al., 2017).

Although the drafts of the genomes of *C. pepo* and other cucurbit species are becoming more complete, little is known about specific gene functions in these crops. Chemically induced plant mutant platforms, such as those generated by EMS, a chemical mutagen that induces single randomly distributed nucleotide changes in DNA (Greene et al., 2003; Sega, 1984), have become an important source of variability for both functional genomic analyses and for plant breeding programs. Once established, the mutant platforms can be used for direct phenotyping, but also for DNA screenings, allowing the detection of single point mutations in a number of specific genes.

High-throughput screenings based on plant phenotyping of large mutant collections are becoming increasingly important as a source of new useful mutants in crop species. The difficulty of direct screenings lies in the large number of plants to be evaluated for each of the characters under study. Since mutations are mostly recessive, the identification of a mutant phenotype requires 8-10 plants to be phenotyped per family, and therefore the screening of 1,000 mutant families requires the assessment of 8,000-10,000 plants. Massive screenings based on seedling early traits have been developed for the detection of agronomic interesting mutants, including mutants with tolerance to biotic and abiotic stresses, as well as insensitive mutants to certain plant hormones or to chemical treatments with specific herbicides or pesticides. For example, the molecular bases of ethylene-regulated processes have been discerned thanks to the analysis of Arabidopsis mutants which are altered in the seedling triple response to ethylene (Bleecker et al., 1988). The response of seedlings to ethylene have been successfully used to screen different EMS mutant collections in Arabidopsis (Guzmán and Ecker, 1990), which allowed the identification of a number of mutants altered in the ethylene pathway (Roman et al., 1995).

Indirect screenings were become more popular since the development of the TILLING (Targeting Induced Local Lesions in Genomes) reverse genetic approach (Colbert et al., 2001). This methodology was firstly applied to detect allelic variants of an specific gene in *Arabidopsis* EMS mutant collections (Greene et al., 2003), but its utilization was rapidly extended to a number of crop species, including rice (Till et al., 2007), maize (Till et al., 2004), barley (Caldwell et al., 2004), *Brassica napus* (Wang et al., 2008), tomato (Minoia et al., 2010), among others. In cucurbit species, TILLING platforms were developed, allowing the detection of new allelic variants for plant breeding in *C. melo* (Dahmani-Mardas et al., 2010; González et al., 2011) and *C. sativus* (Fraenkel et al., 2014; Boualem et al., 2014). An EMS mutant collection has already been published in *C. pepo*, and different target genes were assessed by TILLING (Vicente-Dólera et al., 2014), but the number of mutant squash families is still very poor.

In this paper we present the development and utilization of a new EMS mutant platform in *C. pepo*, made of 3,751 M_2 families. The quality and utility of the collection was confirmed by phenomic screenings of the mutant plants at both seedling and adult stage of development, by using a forward high-throughput screening for ethylene insensitivity, and by assessing the mutation rate and distribution of mutations in two families by whole genome re-sequencing. Results demonstrated that the collection has sufficient quality to address the identification of new *C. pepo* mutants throughout forward genetics, as well as the high density and random distribution of induced mutations in the genome of this crop species for reverse genetics. The established mutant platform constitutes therefore an important resource for functional genomic studies, but also a source of new alleles for squash breeding programs.

3.3. Materials and methods

3.3.1. Generation of the EMS mutant platform

The mutant library was generated in an inbred line derived from the Spanish landrace MUC16. Before starting the generation of the EMS mutant collection, we established the optimal EMS concentration producing the higher mutation density without much altering the seed germination rate, neither the fertility of the M_1 plants. For this end, batches of 200 mature seeds of MUC16 were immersed in bottles containing 0-2% EMS in 200 ml of deionized water and stirred in a shaker for either 12 or 24 h at 22ºC in darkness. After the EMS treatment, seeds were washed twice in 200 ml of 3% $Na_2S_2O_3$ buffer for 30 min at

room temperature with gentle shaking, followed by three washes in 200 ml distilled water for 30 min each. Control seeds were treated with distilled water in the same manner. Treated seeds were germinated in a commercial nursery following standard local practices. After evaluating the germination rate, seedlings were grown under standard greenhouse conditions in Almería (Spain) to obtain mature M_1 plants. Each M_1 plant was then self-pollinated, and M_2 seeds extracted from individual fruits and stored. The fertility of M_1 plants was assessed as the percentage of plants producing more than 20 viable seeds per fruit, i.e. those fruits producing less than 20 seeds were considered to be derived from male- or female-sterile plants. From those experiments we concluded that the optimal mutagenic treatment was that containing 0.3% EMS for a total of 12 h.

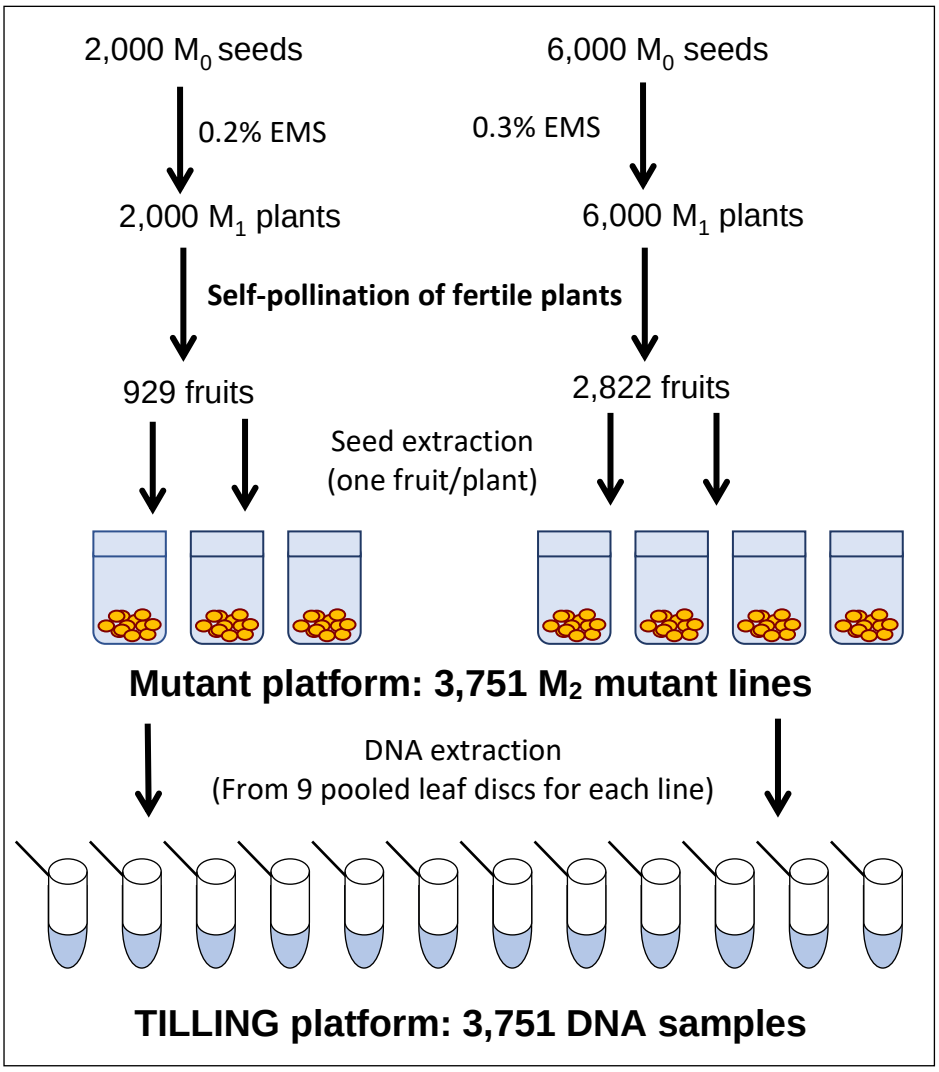


Figure 1. Steps in the construction of the mutant and TILLING platforms of *C. pepo*, each one composed of 3,751 seed and DNA samples, respectively.

To generate the *C. pepo* mutant collection, we treated 6,000 and 2,000 MUC16 seeds with either 0.3% or 0.2% EMS for 12 h, respectively. Those EMS doses did not reduce the germination rate of the MUC16 seeds but lessened the fertility of the plants to nearly 50%. Plants derived from the 6,000 and 2,000 treated seeds were selfed to achieve a total of 2,822 and 929 fruits with more than 20 viable seeds (Figure 1). Consequently, the generated mutant collection comprises a total of 3,751 M_2 families. Regarding the mutant seeds availability, we currently have an average of 160 seeds per M_2 family, and around 600 seeds for each of the 900 M_3 families that were already propagated.

3.3.2. Screening of the mutant library and generation of a TILLING platform

The quality of the mutant collection was firstly tested by evaluating the percentage of easily observable traits in 9 seedlings of each mutant family. A total of 30,008 seedlings were grown, and the alterations in the development and pigmentation of cotyledons and first leaves, in comparison with the background genotype MUC16, recorded for each plant. For each M_2 mutant family we also collected together one leaf disc from eight individual seedlings for DNA extraction, developing so a TILLING platform consisting of 3,751 DNA samples.

The M_2 plants from 284 families in the 0.3% treatment were also grown to maturity under standard greenhouse conditions, determining the percentage of alterations in vegetative and reproductive developmental traits, including plant vigor and height, leaf morphology, root conformation, female and male flower development, sex related traits, and fruit color, size, shape. These M_2 families were also selfed to obtain the M_3 generation.

3.3.3. Assay of triple response to ethylene

To test the quality of the squash mutant collection we also performed a high-throughput screening of the 3,751 M_2 families for mutants in the ethylene-response pathway. The triple response of etiolated seedlings to ethylene was so used to detect M_2 families segregating for ethylene insensitive mutants. Eight seeds of each M_2 family were sown in seedlings trays, germinated for two days in the absence of ethylene and then introduced into a growth chamber containing 50 ppm of ethylene in darkness for a total of 5 days. The ethylene-insensitive seedlings were perfectly distinguishable since they did not reduce the length of their hypocotyls, protruding therefore over the rest of the ethylene-sensitive seedlings in the growth chamber. By using this ethylene test we detected four segregating M_2 families. The ethylene insensitivity of these four families was confirmed by evaluating the response to

ethylene within a higher number of plants in each selected M_2 family. The ethylene insensitive plants were then backcrossed with the inbred line MUC16 for two generations and the resulting progeny (BC_2) selfed to achieve the BC_2S_1 generation. These generation was again evaluated by the ethylene triple response.

Data were analyzed by multiple comparisons by analysis of variance (ANOVA) with significance level $P<0.05$, and each two means were compared with the Fisher's least significant difference (LSD) method. The chi-square analyses were performed using the statistical software Statgraphics Centurion XVI.

3.3.4. Library preparation and sequencing

Whole genome sequencing was performed at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain). Four DNA samples were sequenced for two random M_2 families from the *C. pepo* mutant collection: family 435 corresponding to mutant family 1 (L1), and family 1,717 corresponding to mutant family 2 (L2). Each sample comprises 10 random plants from each family. Paired-end multiplex libraries were prepared according to manufacturer's instructions with KAPA Library Preparation kit (Kapa Biosystems). Libraries were loaded to Illumina flowcells for cluster generation prior to producing 126 base read pairs on a HiSeq2000 instrument following the Illumina protocol. Base calling and quality control were done with the Illumina RTA sequence analysis pipeline according to manufacturer's instructions. Based on the genomic sequence of *Curcubita pepo* (genome v4.1) the average fold effective coverage of all the samples was in the range between 13.18 and 17.90 (Supplementary Table 3.1).

3.3.5. Data analysis

Sequencing reads were trimmed to avoid any remaining adaptors. Resulting reads were mapped to the *C. pepo* genome v4.1 (cucurbitgenomics.org) using the GEM3 toolkit (Marco-Sola et al., 2012) allowing up to 8% mismatches. Alignment files (BAM format) containing only properly paired, uniquely mapping, reads were processed using picard tools version 1.110 (Picard Tools - By Broad Institute) to add read groups and remove duplicates. The Genome Analysis Tool Kit (GATK) (McKenna et al., 2010) version 3.6 was used for local realignment and base recalibration. Processed BAM files were submitted to variant calling for single nucleotide variants and small insertions and deletions using HaplotypeCaller (GATK v3.6). Functional annotations were added to the resulting VCF using SnpEff

(Cingolani et al., 2012) with the gene annotation provided for *C. pepo* genome v4.1 (cucurbitgenomics.org).

3.3.6. Variant filtering

Variants were removed when depth of coverage was less than 8 or greater than 100 in at least one of the samples of each family (L1 and L2). This allowed us not only to exclude positions for which the coverage depth was low, but also positions that might fall in segmental duplications. From the whole genome (263 Mb), after filtering we keep 174 Mb for family L1 and 178 Mb for family L2.

Wrongly mapped reads are difficult to account for and can result in an increased false discovery of genetic variants. In order to remove undetected duplications or repetitive regions that could mislead the alignment program, we further filtered out genomic regions with unusually high number of variants. To identify those regions, a sliding window approach was used, with a window size of 1kb and a step size of 500 bp. All genomic 1 kb windows with more than one variant where discarded. As a result, we discarded 25 Mb of total genomic regions from L1 and 26 Mb from L2.

To determine if a mutation has been caused by EMS in one of the families, it is necessary to have reliable information for that particular position in both families (L1 and L2). After applying the coverage filter removing the hyperpolymorphic regions the intersection of the L1 and L2 genomic positions was computed. The shared reliable regions contained a total of 144.5 Mb.

3.4. Results

3.4.1. Generation of an EMS mutant collection in *C. pepo*

In order to optimize the EMS induced mutagenesis, batches of 200 seeds of MUC16, a *C. pepo* inbred line that was also used for genome sequencing (Montero-Pau et al., 2017), were treated with different concentrations of EMS for 12 and 24 h. The efficiency of the mutagenesis treatments was then estimated by assessing two parameters on M_1 plants: seed germination and plant fertility. Figure 2 shows the reduction in the percentage of germination and in the fertility of M_1 plants subjected to the different treatments.

In seeds treated for 12 h, no significant reduction in germination rates was observed in 0.3-1% EMS treatments, but in 1.5% and 2% EMS treatments the germination was highly reduced (Figure 2). Most of the M_1 plants which were treated for 12 h, exhibited growth

retardation at seedling stages, but all of them recovered and flowered. We found an inverse correlation between mutagen dose and M_1 plant fertility, assessed as the number of selfed fruits, each one having more than 20 viable seeds (Figure 2). At 1.5% and 2% EMS, 70% and 85% of the plants were considered sterile because they either did not set fruit or yielded fruits with less than 20 seeds after selfing.

In EMS treatments for 24 h, the 1.5% and 2% treatment were omitted. Even so, the germination rate was highly reduced, reaching 57% germination after 1% EMS treatment (Figure 2). The fertility of the M_1 plants was greatly reduced in EMS treatments for 24 h, showing less than 75% fertility in the treatments with the different doses (Figure 2). Taken together, the results demonstrated that, even at levels that barely reduced the germination rate, the fertility of plants was strongly affected. Therefore, the most appropriated EMS dosage to produce the mutant collection was selected based on M_1 plant fertility. The 0.3% EMS for 12 h was selected as the most suitable treatment for mutagenesis of MUC16 seeds. This EMS dosage did not affect the germination rate of the M_1 plants, but reduced M_1 plant fertility to about 54% (Figure 2). Fertile plants yielded fruits that contained more than 100 seeds after selfing, therefore not compromising the production of the M_2 generation. In addition, we selected a slightly reduced EMS concentration (0.2%) to obtain one third of the mutant families in our collection.

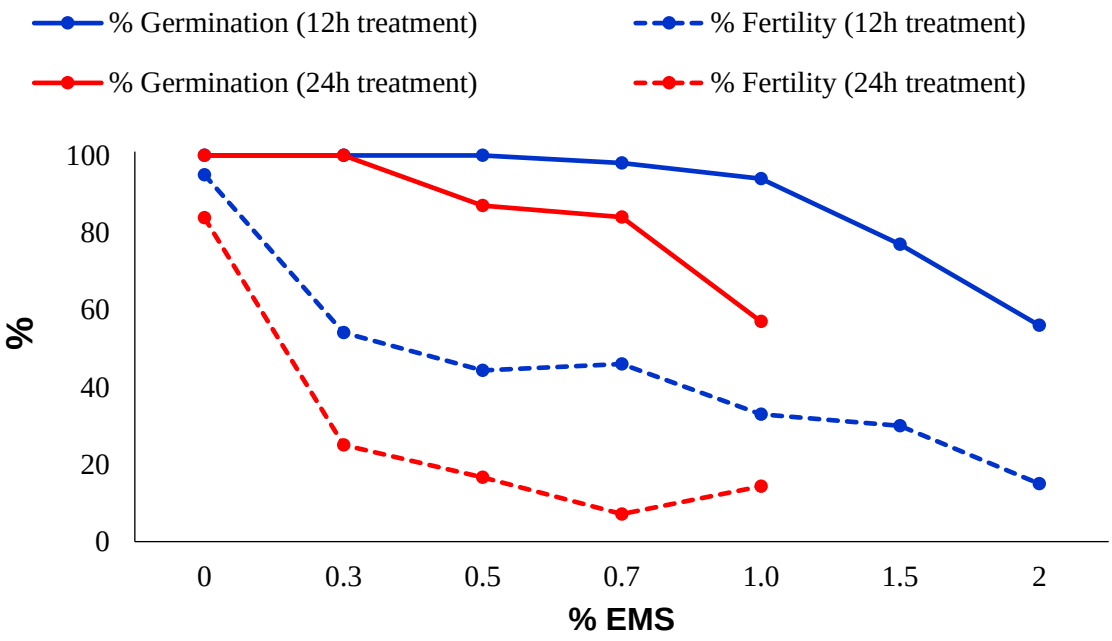


Figure 2. Effect of different EMS treatments on M_1 seed germination and plant fertility.

For generating the final mutant collection, 8,000 MUC16 seeds were mutagenized; 6.000 were treated with 0.3% EMS and 2000 seeds with 0.2% EMS for 12 h at 22ºC. The germinated seed was grown in a greenhouse and each M_1 fertile plant selfed to obtain the M_2 offspring. A total of 3,751 M_2 families were obtained, 929 and 2,822 for the 0.2% and 0.3% doses of EMS, respectively (Figure 1). Nine plants were grown from each mutagenized family, and leaf material from each plant was collected and mixed for DNA extraction. A total of 3,751 samples of DNA were isolated to establish a TILLING platform of the *C. pepo* mutant collection (Figure 1).

3.4.2. Phenotypic variations in the mutant collection

The quality of the mutant collection was firstly tested by evaluating the phenotypes of 9 plants of each M_2 family at the seedling stage of development, focusing on morphological traits that were easily observable in seedlings. The observable changes, which appeared in the cotyledons and leaves of some plants in relation to the genetic background MUC16, were recorded for each M_2 family. Figure 3 shows most of the seedling mutant phenotypes in cotyledons and leaves.

A number of M_2 families showed alterations in the number and development of cotyledons, other in the development of leaves, and others in the pigmentation of vegetative organs. The percentage of families with typical mutations such as albino plants was 0.54% and 1.2% for the families derived from either 0.2% or 0.3% EMS treatments (Table 1).

Table 1. Percentage of phenotypic variations in the seedlings of the 3,751 M_2 mutant families of *C. pepo*.

	0.2% EMS (929 families)		**0.3% EMS (2,822 families)**		**Total (3,751 families)**
	Number of families	Frequency (%)	Number of families	Frequency (%)	Frequency (%)
Albino	5	0.54	33	1.2	1.04
Tri-cotyledon	13	1.4	29	1.03	1.12
Negative gravitropism	3	0.32	36	1.28	1.04
Cotyledon deformation	10	1.08	95	3.4	2.83
Dwarfism	9	0.97	62	2.2	1.89
Variegation	4	0.43	33	1.2	1.01
Depigmentation	18	1.94	50	1.77	1.81
Orange color	0	0	2	0.07	0.05
Ethylene-Insensitive	0	0	4	0.14	0.11
Total M_2 families	**62**	**6.67%**	**244**	**12.19%**	**10.82%**

Other mutant phenotypes, including deformations in cotyledons and leaves, were at higher frequency than albinisms, and the frequency of mutant families with depigmentation in cotyledons and leaves was even higher in those derived from the 0.2% EMS treatment, indicating that density of mutations in the plants derived from this weaker treatment is also elevated (Table 1). By using seedlings at early stages of development, the total number of phenotypic variations in the mutant library was estimated to be 10.82% (Table 1).

Figure 3. Seedling of *C. pepo* M_2 mutant families, showing certain phenotypic alterations: three-cotyledon seedlings, deformation in cotyledons and leaves, variegation and depigmentation in cotyledons and leaves, albino seedlings and dwarfisms.

In adult plants the estimation of phenotypic variations was increased, since we not only phenotyped the vegetative organs of the plant but also flowers and fruits (Figure 4). From a total of 284 M_2 families, vegetative development alterations were observed in 9.15% of the mutant families (Table 2), a similar frequency of vegetative development variations detected in the complete library at the seedling stage. However, in adult plants we also detected different alterations in the morphology of female and male flowers (7.04% and 4.93%, respectively) as well as in fruit color, size and shape (7.04%). The total frequency of phenotypic variations was increased to 28.17% which represents a high frequency of variation for forward genetic approach.

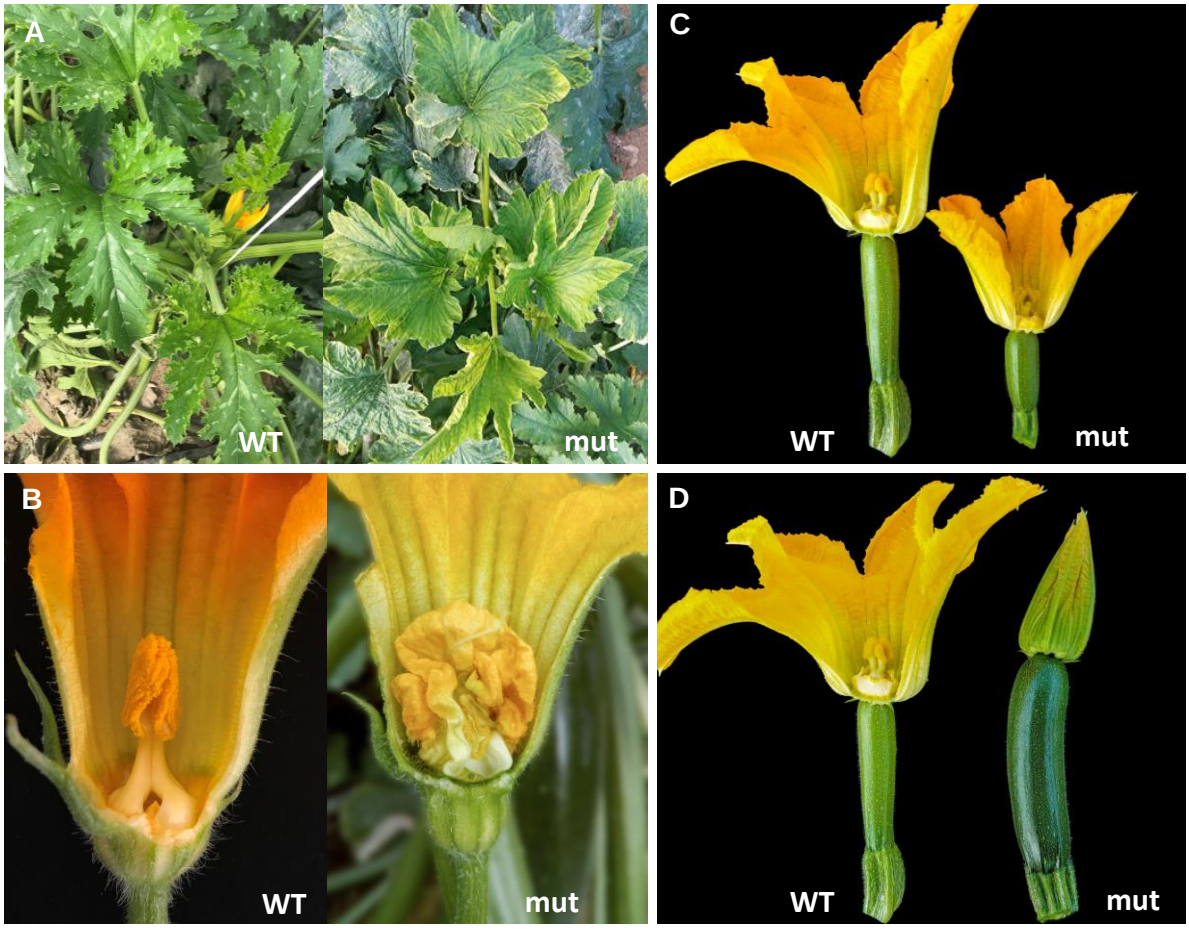


Figure 4. **Mutant phenotypes observed in M_2 mutant families.** (**A**) Alterations in leaf morphology. (**B**) Alterations in male flower development. (**C**) Alterations in female flower size, and ovary shape and size. (**D**) Alterations in fruit set and female flower development (the mutant fruit has a parthenocarpic development).

From a total of 284 M_2 families, vegetative development alterations were observed in 9.15% of the mutant families (Table 2), a similar frequency of vegetative development variations

was detected in the complete library at the seedling stage. However, in adult plants we also detected different alterations in the morphology of female and male flowers (7.04% and 4.93%, respectively) as well as in fruit color, size and shape (7.04%). The total frequency of phenotypic variations was increased to 28.17% which represents a high frequency of variation for forward genetic approach.

Table 2. Percentage of phenotypic variations in adult plants from 284 M_2 families derived from the 0.3% EMS treatment.

Phenotypic variation	Number of families	Percentage (%)
Vegetative development	26	9.15
Male flower	14	4.93
Female flowers	20	7.04
Fruit set and development	20	7.04
Total families affected	**80**	**28.17%**

3.4.3. High-throughput screening of the collection for ethylene insensitivity

To test the quality of the *C. pepo* mutant library for forward genetic analysis, we performed a high-throughput screening for ethylene insensitivity by using the triple response of etiolated seedlings to ethylene. Nine seeds of each family were germinated and treated with gaseous ethylene for 48 h. In the WT genotypic background (line MUC16) as well as in the vast majority of the M_2 families, all plants showed a positive triple response to ethylene. Seedlings displayed a decrease in the length of hypocotyl and roots, an increased hypocotyl thickness and greater curvature of the apical hook, characteristic that were rapidly observed in the germinating seeds (Figure 5). Four out of 3,751 families showed a reduced response to ethylene treatment, indicating that some of the plants were insensitive to this hormone. The identified mutants, designated *ein1*, *ein2*, *ein3* and *EIN4*, developed larger and thinner hypocotyls, as well as lager roots in comparison with WT plants of the same family (Figure 5). In the *ein1*, *ein2* and *ein3* families, some of plants showed an intermediate phenotype between WT and *ein* mutants.

From a total of 284 M_2 families, vegetative development alterations were observed in 9.15% of the mutant families (Table 2), a similar frequency of vegetative development variations was detected in the complete library at the seedling stage. However, in adult plants we also detected different alterations in the morphology of female and male flowers (7.04% and

4.93%, respectively) as well as in fruit color, size and shape (7.04%). The total frequency of phenotypic variations was increased to 28.17% which represents a high frequency of variation for forward genetic approach.

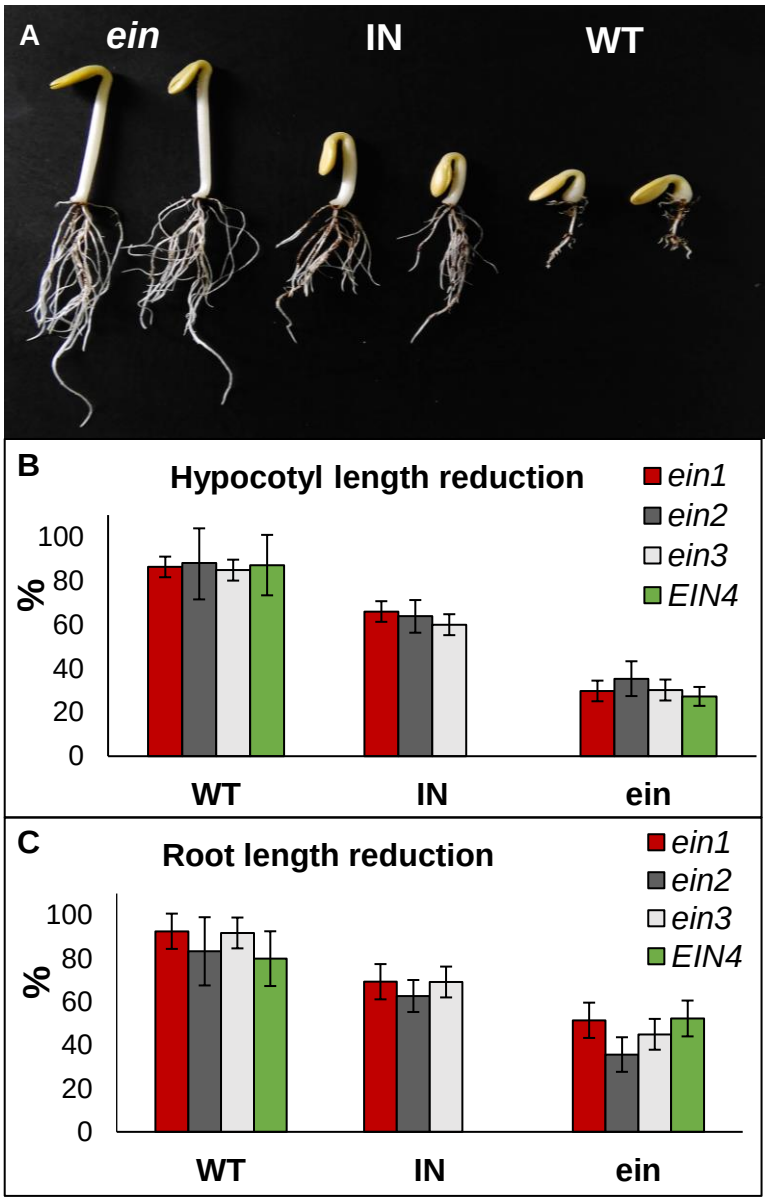


Figure 5. (**A**) Ethylene triple response phenotypes in etiolated seedlings of the *ein1* mutant family of *C. pepo*. WT plants showed a positive triple response, characterized by a reduction in the length of hypocotyl and root, and a thicker hypocotyl. The ethylene insensitive (*ein*) plants do not respond to ethylene and show longer and thinner hypocotyl, and longer roots. Certain plants in the family showed an intermediate (IN) triple response phenotype. (**B**, **C**) Percentage of reduction of hypocotyl and root length of ethylene treated WT and e*in* mutants in comparison with untreated seedlings of the background genotype MUC16. Errors bars represent standard errors. Note that in the *ein1-3* mutant families there is a distinguishable intermediate (IN) phenotype to ethylene response.

Only 1-4 plants out of nine analyzed seeds in each mutant family showed an ethylene insensitive phenotype. To confirm the ethylene insensitivity of these mutant families, we increased the number of plants in segregating progenies of each family and reduced the number of other mutations in each of the families by backcross. The M_2 mutant plants were backcrossed with the background WT genotype MUC16 for two generations, and the BC_2 progeny selfed to obtain the BC_2S_1 generation. The backcrossed plants had the same ethylene insensitive phenotype as the original M_2 plants. Segregations also demonstrated that the four mutations were semi-dominant, and that intermediate phenotype for ethylene triple response corresponded to heterozygous plants for the mutations (Table 3).

Table 3. Segregation of ethylene insensitive mutants in the backcrossing (BC_1 and BC_2), and the first selfed (BC_2S_1) generations.

		Number of plants					
Mutant family	**Generation**	**WT**	**INT**	***ein***	**Expected segregation**	**χ^2**	***P*-value**
	BC_1	-	62	-		-	-
ein1	BC_2	20	24	-	1:1	0.36	0.54
	BC_2S_1	116	187	115	1:2:1	4.64	0.1
	BC_1	-	79	-		-	-
ein2	BC_2	42	41	-	1:1	0.01	0.91
	BC_2S_1	129	237	119	1:2:1	0.66	0.72
	BC_1	-	42	-		-	-
ein3	BC_2	31	27	-	1:1	0.28	0.6
	BC_2S_1	98	178	86	1:2:1	0.89	0.64
	BC_1	34	-	30	1:1	0.25	0.62
EIN4	BC_2	120	-	125	1:1	0.1	0.75

Plants were classified according to their triple response to ethylene: WT, ethylene-sensitive seedlings; INT, intermediate response to ethylene; *ein*, ethylene-insensitive seedlings. *EIN4* plants only produce male flowers and were unable to be selfed.

Furthermore, by using a higher number of plants for each mutant family, we assessed the reduction in the length of hypocotyls and roots of ethylene treated seeds in comparison with untreated seeds of the background genotype MUC16 (Figure 5). In the segregating progenies of *ein1*, *ein2* and *ein3* mutants, we distinguished three phenotypes, although only two phenotypes were found in the *EIN4* family (Figure 5).

3.4.4. Density and spectrum of EMS mutations

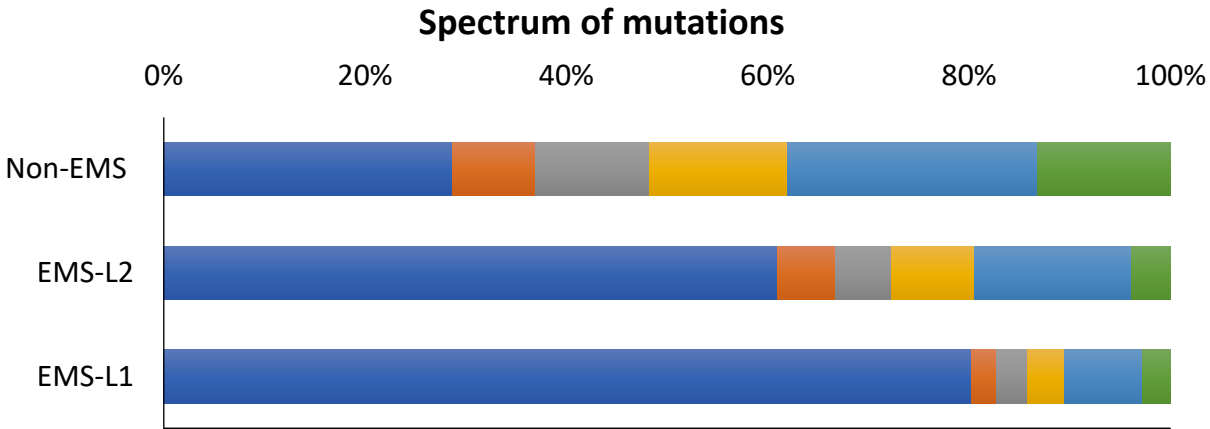
Next generation sequencing (NGS) technology makes whole-genome sequencing a practical method for mutation identification and mapping. DNA from 40 plants in segregating progenies of two mutant families (L1 and L2) was used for WGS. Since WGS was made simultaneously with the phenotyping of seedlings and plants, these two families were randomly selected from the collection. They were later phenotyped, but no alterations were detected in easily observable morphological traits. Four samples, each one containing a pooled DNA from at least 10 plants in segregating progenies, were sequenced for each family. Sequencing data from the eight samples showed a mean coverage of 92.4% and 95.5% of the whole genome for L1 and L2, respectively. After variant calling for each individual sample, SNVs with GQ>20 were firstly filtered for a minimum coverage threshold of 8 reads in at least one sample per family, and maximum coverage threshold of 100 reads (median coverage per sample was 35 reads). The regions with higher coverage were filtered out because their high number of SNVs is most likely caused by wrong mapped interspersed DNA sequences. This filtered out 15% of the bases, resulting in a genome coverage of 67% and 67.8% for families L1 and L2, respectively (Supplementary Table 3.4).

A second filtering step was based on the density of variants, filtering out those SNVs that had another variant closer than 1,000 bp because they were likely caused by sequencing or mapping errors. That reduced the coverage of the reference genome to about 60% in the two families, but also reduced the number of variants to 6,140 in L1 and 5,252 in L2. Most of these variants (4,031) were common to both L1 and L2 (Supplementary Table 3.4), likely representing spontaneous nucleotide polymorphisms in MUC16 genetic background. They account for 65.7% and 76.8% of SNVs in L1 and L2, respectively. Those that were homozygous in both L1 and L2 (3,136) were probably fixed during the 2-3 extra rounds of selfing that MUC16 reference genotype was subjected before EMS treatments. The other shared SNVs could be spontaneous polymorphisms still segregating in one or the two families. Given that the MUC16 line has more than nine generations of selfing and taking into account that *C. pepo* genome is duplicated (Montero-Pau et al., 2017), we cannot rule out that these shared SNVs were the result of mapping errors in reads coming from duplicated genomic regions.

The specific variants sum a total of 2,209 in L1 and 1,226 in L2 from a total of 144.5 Mb (55.4% of genome) shared by the two families after filtering (Supplementary Table 3.4). Given the high percentage of spontaneous mutations among SNVs, the homozygous family-

specific SNVs would mostly be spontaneous polymorphisms that were fixed for the alternative allele in each one of the families after mutagenesis. Therefore, we discarded these family-specific homozygous SNVs as truly EMS mutations. The family-specific heterozygous SNVs (1,704 and 859 in L1 and L2, respectively) were therefore selected as the tangible EMS induced DNA mutations, which reveals a mutation density of 11.8 mutation/Mb (1 mutation/85 kb) in L1, and 6 mutations/Mb (1 mutation/167 kb) in L2.

The most frequent mutations were the EMS canonical transitions GC to AT (80.3% in L1 and 61% in L2), followed by TA to CG transitions (7% and 15.6%) (Figure 6). The frequency of transversion mutations (GC>CG, GC>TA, TA>AT and TA>GC) was relatively low and summarizes 12.7% in L1 and 23.4% in L2 (Figure 6). The transition to transversion ratio was therefore 6.9% and 3.3% in L1 and L2, respectively. The spectrum of mutations in the fraction of SNVs that were discarded as non-EMS (those that were shared by the two families, and those that were homozygous in only one of the families), was not enriched in GC to AT transitions (Figure 6), suggesting that they were not likely induced by the mutagen but were polymorphism of the background genotype.



	EMS-L1	EMS-L2	Non-EMS
GC>AT	1368	524	1158
GC>CG	41	49	331
GC>TA	54	48	457
TA>AT	62	71	551
TA>CG	132	134	1001
TA>GC	47	33	533

Figure 6. Spectrum and frequency of and EMS and non-EMS mutations in L1 and L2 families. EMS mutations correspond to those that were heterozygous and specific for one of the mutant family. Homozygous family specific SNVs and shared SNVs between L1 and L2 families were considered non-EMS.

To investigate sequence biases in EMS-induced mutations, we analyzed the frequency of preferential bases flanking ±2 positions around the most frequent transitions (GC to AT) respect to the genome background (Figure 7). In the background genome, the two bases surrounding a C, were more frequently A and T than C and G. In -1 and +1 positions of the most numerous GC to AT transitions, pyrimidines are more frequent than expected, with significant biases to T in the -1 position and to T and C in the +1 position. No bias was observed in the +2 position, but an excess of Cs occurred at the -2. In comparison with the background genome, these nucleotide biases were not detected around the GC to AT spontaneous SNVs observed in the mutant families (Figure 7).

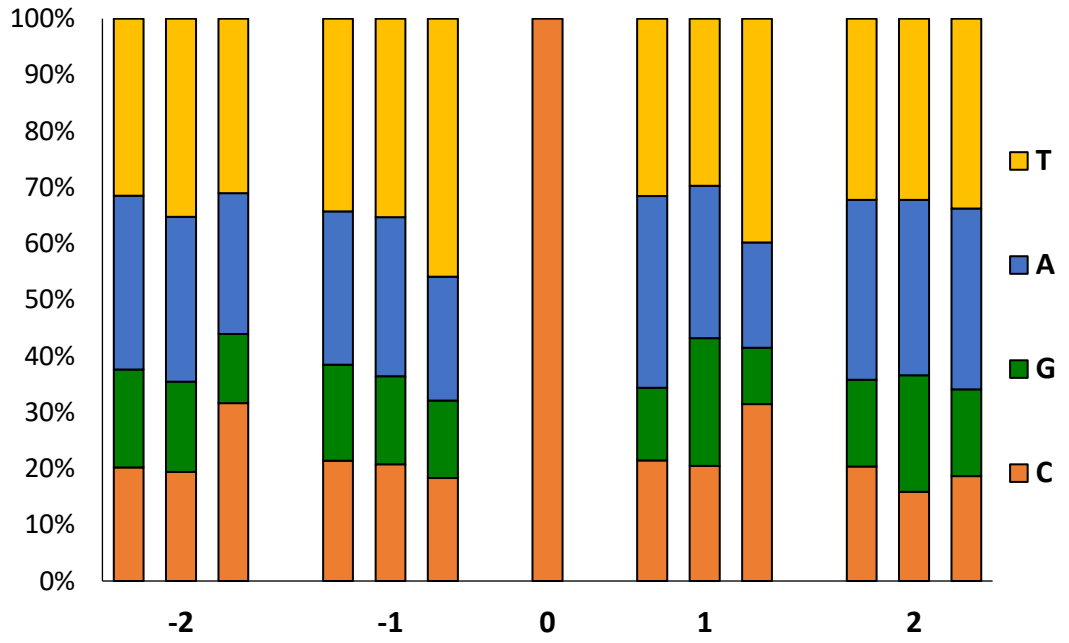


Figure 7. Frequency of the sequences flanking EMS-induced mutations. At each position (-2, -1, +1 and +2), left bars correspond to the frequency of bases flanking C in the background genome, the middle bars, to flanking sites of the non-EMS CG>TA variants (natural polymorphisms shared by the two families respect to the reference genome), and the right bars to the canonical EMS mutations (CG>AT).

3.4.5. Distribution and impact of EMS mutations

The distribution of EMS mutations among the 20 chromosomes of *C. pepo* is shown in Figure 8. The mutations were evenly distributed throughout the genome with a density ranging from 5 to 11.7 mutations/Mb in chromosomes 15 and 14 of L1, and 1.5 to 13.6 mutations/Mb on chromosomes 7 and 6 of L2. No preferential distribution was detected among euchromatic and heterochromatic regions in the different chromosomes. The

causative effects of the mutation on gene function was studied by using the SnpEff program. On the base of *C. pepo* genome annotation, many of the EMS mutations were located in regions with low predicted impact on gene function, including intergenic regions (23.6%), downstream and upstream (55%) regions and introns (11.1%).

A reduced number of EMS mutations affected, however, to exons (7.3%), 5'- and 3'-UTR regions (2.4%) and splice sites (0.5%), which could likely alter gene function (Table 4). Of the 422 exonic mutations 281 (66.6%) were missense mutations, 11 (2.6%) nonsense mutations, and 130 (30.8%) silent (Table 4). Regarding their impact, 0.3% of the detected EMS mutations had a predicted high- (e.g. nonsense, frameshift, and splice acceptor or donor mutations), 4.8% a moderate- (e.g. missense mutations), 92.1% a modifier- (e.g. intron and intergenic mutations), and 2.8% a low-impact (e.g. synonymous mutations) on gene function.

Table 4. Location and functional impact of EMS mutations in L1 and L2 mutant families of *C. pepo*.

	Mutation	**L1**	**L2**	**Total**	**%**
Gene region	Exon	311	111	422	7.3
	Intron	479	166	645	11.1
	Splice sites	23	6	29	0.5
	Intergenic	821	548	1,369	23.6
	5´- and 3´- UTR	101	41	142	2.4
	Downstream and upstream	2,202	990	3,192	55
Protein change	Missense	207	74	281	66.6
	Nonsense	9	2	11	2.6
	Silent	95	35	130	30.8
Impact	Low-impact mutations	117	40	157	3.5
	Moderate-impact mutations	207	74	281	6.3
	Modifier-impact mutations	2,776	1,197	3,973	89.7
	High-impact mutations	16	3	19	0.4

Predicted low-impact mutations correspond to synonymous and splice region variants; Moderate-impact mutations, to missense variants, and high-impact mutations, to stop gained, and splice acceptor and splice donor variants (Supplementary Tables 3.2 and 3.3).

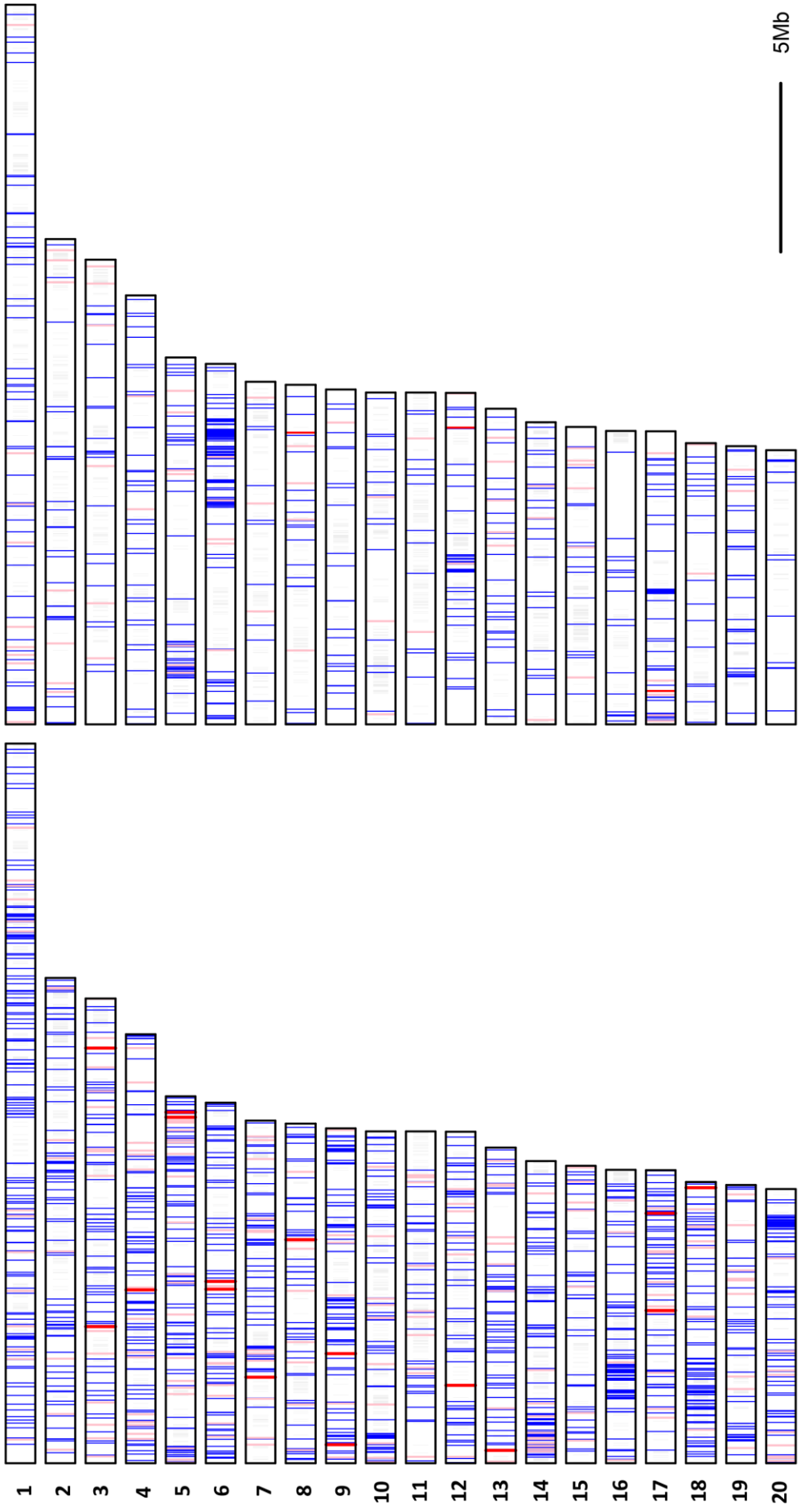


Figure 8. Chromosome distribution (1-20) of EMS-induced mutations in L1 and L2 mutant families. Depending on the predicted impact, EMS-induced variants are plotted as red (high impact), pink (moderate) or blue (low or very low impact). Regions discarded because of the high density of variants are shaded in grey.

Of the 19 high-impact mutations, 10 were nonsense, and 3 and 5 occurred on splice acceptor and splice donor sites of target genes, respectively (Table 4). The genes affected by a premature stop codon, which are likely those with higher impact on gene function, code for proteins like: E3 ubiquitin-ligase PRT6 and probable RNA helicase SDE3 in family L2, and for chloroplastic glyceraldehyde-3-phosphate dehydrogenase; multi-bridging factor 1a-like; probable LRR receptor like serine threonine- kinase At4g26540; low-temperature-induced cysteinase; nuclear transcription factor Y subunit B-9; ADP-ribosylation factor-related 1; Protease 2 and aquaporin TIP2-1 in family L1 (Supplementary Tables 3.2 and 3.3).

3.5. Discussion

3.5.1. Quality and usefulness of the *C. pepo* EMS mutant library

A mutant platform has been established in the crop species *C. pepo*, consisting of 3,751 M_2 families with an average mutation density of 8.9 mutations/Mb (1 mutation per 112 kb). The artificial mutagenesis was found to generate more than 10% of morphological variants in early seedling developmental stages, and more than 30% of morphological alterations in adult plants. This mutation density, detected by both phenomic and genomic approaches, is quite similar or higher to that observed in other plant EMS collections (Boualem et al., 2014; Dahmani-Mardas et al., 2010; González et al., 2011; Vicente-Dólera et al., 2014), indicating that the library will be a high-quality tool for both forward and reverse genetic analyses.

The efficiency of chemical mutagenesis is variable, depending not only on the nature and dosage of the mutagen, but also on the considered species and the background genotype (Chen et al., 2014). The *C. pepo* library was generated with a lower EMS dosage than in other plant species, including the related cucurbit species. *C. melo* collection was generated by 1.5% EMS, while that of *C. sativus* was generated by using 0.5 and 0.75% EMS (Boualem et al., 2014; Dahmani-Mardas et al., 2010). Similar doses (0.7-1.5%) were used in tomato (Minoia et al., 2010), maize (Till et al., 2004) and rice (Till et al., 2007). Despite the use of lower EMS dosage, the *C. pepo* library did not reduce the mutation density, suggesting that *C. pepo* is more sensitive to EMS than other species. Seed germination was similarly affected in the different species, but the more harmful effect of EMS treatment in squash was observed on plant fertility. Consequently, the selection of EMS dose could not be made on the basis of seed germination but based on M_1 plant fertility. At the end an EMS dosage was utilized (0.2% and 0.3%) that resulted in 100% of M_1 seed survival, but reduced fertility to about 50%, and allowed therefore the production of large populations of viable M_2 seeds.

The alterations found in seedlings morphological traits, which is a good indicator of a high-quality mutant collection (Wang et al., 2008), reached 10.82%, is similar to that reported for other well characterized mutant collections. Thus, the percentage of M_2 families with albino and chlorophyll deficient families in our collection (1.85%) was very similar to the 1.3% rate found in a previous squash library (Vicente-Dólera et al., 2014), as well as the 0.6% and 2.1% rates observed in the mutant collection of *C. sativus* (Boualem et al., 2014), and muskmelon (González et al., 2011), respectively. Similar frequency of albino mutants, were found in tomato (1.7%; (Minoia et al., 2010)) and in common bean (1.53%; (Porch et al., 2009)). The frequency of the dwarf phenotype in our library (1.89%) was also similar to that observed in other species, including tomato, squash and soybean (Minoia et al., 2010; Tsuda et al., 2015; Vicente-Dólera et al., 2014). These data confirm the efficiency of the mutagenesis and demonstrate that the generated mutant library has a high density of mutations, where potential mutant phenotypes can be certainly found by a forward genetic approach.

The frequency of mutant phenotypes in adult plants (28.17%) further encourages the use of our library in forward genetic analyses. Many of them were variations in vegetative developmental traits such as leave shape and color, and growth habit, but many of the mutant families were altered in male and female flower development (11.97%), including sex determination and sex expression mechanisms (Table 2). The high-throughput screening for ethylene insensitivity by using the triple response validate the use of this mutant library in forward genetic approach. A high-throughput phenomic screening with more than 35,000 plants resulted in four ethylene-insensitive mutants that showed a reduced response to ethylene in root length and hypocotyl length and thickness. This hormone controls flower and fruit development (Martínez et al., 2013), sex expression and sex determination (Manzano et al., 2010a, 2011, 2013), and chilling injury in refrigerated zucchini fruit (Megías et al., 2016). Therefore, these mutants could be of value not only for investigating the function of ethylene receptor and signaling genes in *C. pepo*, but also for breeding a number of agronomic traits in this and other *Cucurbita* crop species, providing they can be crossed with *C. pepo*. Moreover, knowledge of gene function in *C. pepo* would permit the biotechnological manipulation of other plant species through transgenesis or genome-editing technologies.

3.5.2. Density, spectrum, distribution and impact of EMS mutations

The mutation density in the squash mutant platform, estimated by whole-genome sequencing (WGS) of two independent mutant families, was high if compared with other libraries. In addition to sequence quality and coverage, we also filtered SNVs that likely resulted from sequencing and mapping errors (regions with excessive coverage and with high mutation density, probably caused by wrong mapping of repetitive DNA reads), and spontaneous point mutations (SNVs shared by the two families). Discarding clustered SNVs (those that were separated less than 1000 bp) is also in accordance with a study in *Drosophila*, where it was determined that clustered SNPs, which in this case were less than 500 bp apart from one another, were actually natural polymorphisms that arose from gene conversion events (Blumenstiel et al., 2009). After filtering, a total of 2,563 EMS mutations were identified in both L1 and L2 families. The average density was estimated as 1 mutation/111 kb in 144.5 Mb of scored genome, which is slightly higher respect to that so far available and published for a smaller squash mutant population of 768 M_2/M_3 families (1/133 kb; (Vicente-Dólera et al., 2014)). The estimated mutation density was also higher in comparison to that published for other EMS mutant collections in the related cucurbit species melon (1 mutation/573 kb; (Dahmani-Mardas et al., 2010)) and cucumber (1 mutation/1147 kb; (Boualem et al., 2014)), although in all these cases, the estimations were based on the TILLING of specific target genes and not on WGS. The higher mutation density observed in squash in comparison with other EMS families in related cucurbit species, could be related with the whole-genome duplication event observed *Cucurbita* species (Montero-Pau et al., 2017), which could help in withstanding the mutagen action, as appears to occur in tetraploid and hexaploid wheat, where observed EMS mutation density reaches 1/40 kb and 1/24 kb, respectively (Slade et al., 2005). The twofold differences in mutation density observed between the two sequenced families has been also reported in mutant libraries of other species like soybean (Tsuda et al., 2015), tomato (Shirasawa et al., 2016), and C*aenorhabditis elegans* (Sarin et al., 2010), where differences among certain families reached up to tenfold.

The most frequent EMS-induced mutations found in the mutant collection were GC to AT transitions, which are known to result from the alkylation of guanine residues. In other mutant libraries of *Arabidopsis*, maize and pea, the frequency of GC to AT transitions represent more than 99% of the EMS mutations (Dalmais et al., 2008; Greene et al., 2003; Till et al., 2004). Whole-genome re-sequencing allowed the identification of 80.3% and 61%

of GC>AT transitions in L1 and L2, respectively, but other less frequent transitions (TA>CG) and transversions (TA>AT, GC>CG, TA>GC and GC>TA) were also identified in the EMS-induced fraction. Although one could think that these non-canonical mutations are false positives, their occurrence and proportion is consistent with that found in other studies in *Drosophila* (Blumenstiel et al., 2009; Cooper et al., 2008), tomato (Minoia et al., 2010; Shirasawa et al., 2016), soybean (Cooper et al., 2008), rice (Till et al., 2007) and barley (Caldwell et al., 2004), suggesting that, although caused by unknown mechanisms, they are also produced by EMS.

The spectrum of spontaneous mutations occurring in the genome of *C. pepo* was different to that generated by EMS. In this fraction, identified as SNVs shared by the two families under analysis or homozygous for one of the families, the percentage of transitions GC to AT (28.6%) and TA to CG (24.8%) was very similar, while the percentage of transversions reached 46.4%. Moreover, the transition to transversion ratio of spontaneous mutations (1.15) is very different to that in the EMS-induced mutations (6.9 and 3.3 in L1 and L2, respectively), which also strengthen the filtering process followed for the selection of EMS-induced mutations in our study.

The artificial mutations were evenly dispersed throughout the 20 chromosomes of *C. pepo*, and occur randomly in the genome, independently of their location in euchromatic or heterocromatic regions. However, EMS action does not appear to be randomly occurring but is prone to happen in some specific DNA sequences. We found that the two contiguous bases at both sides of the base substitution GC to AT transitions were biased respect to the genome background, with an excess of pyrimidine at -1 and +1 positions, and an excess of C at the -2 position (Figure 7). These data are in accordance to that found in other plant species, including tomato (Shirasawa et al., 2016) and Arabidopsis (Greene et al., 2003).

3.6. Conclusions

A large mutation platform has been developed in *C. pepo*, containing 3,751 independent families with an average mutation frequency of 1/111 kb. The mutations are evenly distributed along the complete genome and consist of more than 75% of GC to AT transitions, which are the canonical mutations generated by EMS. The putative impact of the mutation detected by WGS, and the phenotypic variation observed in the new squash-mutant platform corroborate its usefulness in high-throughput mutation discovery for both gene function analysis and plant breeding. The high-throughput screening of the whole library by

using the triple response to ethylene has yielded a total of 4 ethylene insensitive mutants that are presently being analyzed, and other screenings for biotic and abiotic stresses are underway with the aim to detect novel alleles and new gene functions.

4. The ethylene receptors *CpETR1A* and *CpETR2B* cooperate in the control of sex determination in *Cucurbita pepo*

4.1. Abstract

High-throughput screening of an ethyl methanesulfonate-generated mutant collection of *Cucurbita pepo* using the ethylene triple-response test resulted in the identification of two semi-dominant ethylene-insensitive mutants: *etr1a* and *etr2b*. Both mutations altered sex determination mechanisms, promoting conversion of female into bisexual or hermaphrodite flowers, and monoecy into andromonoecy, thereby delaying the transition to female flowering and reducing the number of pistillate flowers per plant. The mutations also altered the growth rate and maturity level of petals and carpels in pistillate flowers, lengthening the time required for flowers to reach anthesis, as well as stimulating the growth rate of ovaries and the parthenocarpic development of fruits. Whole-genome sequencing allowed identification of the causal mutation of the phenotypes, as two missense mutations in the coding region of *CpETR1A* and *CpETR2B*, each one corresponding to one of the duplicates of ethylene receptor genes highly homologous to Arabidopsis *ETR1* and *ETR2*. The phenotypes of homozygous and heterozygous single- and double-mutant plants indicated that the two ethylene receptors cooperate in the control of the ethylene response. The level of ethylene insensitivity, which was determined by the strength of each mutant allele and the dose of wild-type and mutant *etr1a* and *etr2b* alleles, correlated with the degree of phenotypic changes in the mutants.

Key words: *Cucurbita pepo*, ethylene, ethyl methanesulfonate, mutants, sex determination, fruit set.

4.2. Introduction

Monoecious and dioecious plant species produce unisexual flowers (male or female) either in the same plant (monoecy) or in separate plants (dioecy). They are believed to be derived from a hermaphrodite ancestor, by different mechanisms, which result in the suppression of either stamen or carpel primordia development during the formation of female or male flowers, respectively (Dellaporta and Calderon-Urrea, 1993; Pannell, 2017). The genetic control mechanisms underlying sex determination in plants are diverse, ranging from heteromorphic sex chromosomes, as occurs in the dioecious species *Silene latifolia* and *Rumex acetosa*, to a number of non-linked genes, as occurs in the monoecious species of the family *Cucurbitaceae* (Jamilena et al., 2008; Pannell, 2017)

Cultivars of *Cucurbitaceae* species, including *Cucumis sativus* (cucumber), *Cucumis melo* (melon), *Citrullus lanatus* (watermelon) and the species of the genus *Cucurbita* (pumpkins and many squashes), are mainly monoecious, although certain cultivars of *C. sativus*, *C. melo* and *C. lanatus* are andromonoecious, producing male and hermaphrodite flowers on the same plant (Boualem et al., 2009; Malepszy and Niemirowicz-Szczytt, 1991; Manzano et al., 2016; Perl-Treves, 2004). Both monoecious and andromonoecious plants go through two flowering phases of development: an initial phase in which the plant produces only male flowers, and a second phase in which the plant alternates the production of pistillate and male flowers (Manzano et al., 2013, 2014; Perl-Treves, 2004; Zhang et al., 2017). The transition to pistillate flowering, and the number of pistillate flowers per plant, vary within the different cultivars of each species. In *Cucumis*, natural genetic variation includes other sexual phenotypes, such as gynoecious (a plant producing only female flowers) and androecious (a plant producing only male flowers). Neither the andromonoecious phenotype nor the gynoecious and androecious phenotypes have been observed in the genus *Cucurbita*, although some cultivars show a partially andromonoecious phenotype characterized by the occurrence of male and bisexual flowers, that is, pistillate flowers with partially developed stamens and no pollen (Martínez et al., 2014).

Sex determination mechanisms in cucurbit species are controlled by ethylene. Under treatments that reduce ethylene biosynthesis or perception, the monoecious plants are converted into partially or completely andromonoecious ones, demonstrating that ethylene participates in the sex identity of female flowers, and that an ethylene threshold is required to arrest stamen development in the female flower (Manzano *et al*., 2011; Zhang *et al*., 2017). Besides the control of individual floral buds, ethylene also participates in the control of sex expression within the plant. A reduction in ethylene biosynthesis or perception delays the transition to pistillate flowering and reduces the number of pistillate flowers per plant in *C. sativus*, *C. melo*, and *C. pepo*, but has the opposite effect in *C. lanatus* (Manzano et al., 2011, 2014). By contrast, treatment with ethylene or ethylene-releasing agents induces the production of male flowers in *C. lanatus* but promotes the production of pistillate flowers in the other species (Rudich *et al.*, 1969; Byers *et al.*, 1972; Den Nijs and Visser, 1980).

The genes and mutations responsible for cucurbit sex phenotypes are currently being sought. The arrest of stamen development in the female flowers of the different species requires the functioning of the ethylene biosynthesis orthologs *CmACS7, CsACS2, CpACS27A,* and *CitACS4*, which are specifically expressed in the female flowers of *C. melo*, *C. sativus*, *C.*

pepo, and *C. lanatus*, respectively. Loss-of-function mutations in these genes led to andromonoecy in *C. melo*, *C. sativus*, and *C. lanatus* (Boualem et al., 2008, 2009; Ji et al., 2016; Manzano et al., 2016), but to only partial andromonoecy in *C. pepo* (Martínez et al., 2014). The androecious and gynoecious phenotypes also resulted from two independent mutations: in *C. melo*, androecy resulted from a mutation in the *CmACS11* gene (Boualem et al., 2015), while gynoecy was produced by a mutation in the *CmWIP1* gene (Martin et al., 2009). *CmACS11* represses the expression of *CmWIP1* to permit the coexistence of male and female flowers in monoecious species (Boualem et al., 2015).

Having performed extensive screening of an ethyl methanesulfonate (EMS)-generated mutant collection of *C. pepo* in the search for ethylene-insensitive mutants (García et al., 2018) in order to gain further insights into the genetic network regulating sex determination in cucurbits, in this paper we present a molecular and functional characterization of two semi-dominant mutations that affect two ethylene receptors of *C. pepo*: *CpETR1A* and *CpETR2B*. These mutations confer ethylene insensitivity on the plant, resulting in the conversion of female flowers to bisexual or hermaphrodite ones; that is, monoecy to andromonoecy. The mutations also alter the development and maturity of different floral organs in pistillate flowers, including ovaries and fruit.

4.3. Materials and methods

4.3.1. Plant material

The ethylene-insensitive mutants analyzed in this study were selected from a high-throughput screening of a *C. pepo* mutant collection by the triple-response assay (García et al., 2018), consisting of shortening and thickening of hypocotyls and roots in seedlings germinated in the dark with an external input of ethylene (Bleecker et al., 1988). The *etr1a* and *etr2b* mutants analyzed in this paper correspond to the *ein2* and *ein3* mutants isolated by García *et al.* (2018).

Before phenotyping, *etr* mutant plants from each family were crossed for two generations with the background genotype MUC16, and the resulting BC_2 generation was selfed to obtain the BC_2S_1 generation. Given that homozygous *etr1a* and *etr2b* mutants were female sterile, they were always derived from selfed progenies of BC heterozygous plants. Sterility also prevented us from obtaining double homozygous mutants. The heterozygous double mutants (wt/*etr1a* wt/*etr2b*) were obtained by crossing heterozygous wt/*etr1a* as female and homozygous *etr2b/etr2b* as male, and genotyping the offspring for the causal mutations. The

double heterozygous plants were also female sterile, preventing us from obtaining the double homozygous *etr1a/etr1a etr2b/etr2b.*

4.3.2. Phenotyping for sex expression and sex determination traits

First, BC_1S_1 or BC_2S_1 plants from *etr1a* and *etr2b* mutant families were classified according to their level of triple response to ethylene in wild-type (WT), intermediate (wt/*etr*) and ethylene-insensitive mutants *(etr/etr)*, and then transplanted to a greenhouse and grown to maturity under standard crop management, in Almería, Spain. For the ethylene response assay, seeds were germinated for 2 days in the absence of ethylene and then placed in a growth chamber containing 50 ppm ethylene in darkness for 5 days. The identified mutants showed more elongated hypocotyls and roots than WT, resembling seedlings grown in air. The ethylene sensitivity of each mutant genotype was estimated by using three replicates with at least 20 seedlings of the same genotype. Ethylene sensitivity was assessed as the percentage of reduction in hypocotyl length relative to air-grown seedlings: $[(H_0-H_1/H_0)\times100]$, where H_0 corresponds to the hypocotyl length of air-grown seedlings, and H_1 to the hypocotyl length of ethylene-treated seedlings. The final ethylene sensitivity was relativized to the WT seedling response, considering that the WT has 100% ethylene sensitivity, and ethylene insensitivity was calculated as (100–ethylene sensitivity).

The sex phenotype of each plant was determined according to the sex of the flowers in the first 40 nodes of each plant. A minimum of 30 WT, 30 wt/*etr*, and 30 *etr* plants were phenotyped for each mutant family. The sex expression of each genotype was assessed by determining the node at which plants transitioned to pistillate flowering, and the number of male or pistillate flower nodes. The sex phenotype of each individual pistillate flower was assessed by the so-called andromonoecy index (AI) (Manzano et al., 2016; Martínez et al., 2014). Pistillate flowers were separated into three phenotypic classes that were given a score from 1 to 3 according to the degree of their stamen development: female (AI=1), showing no stamen development; bisexual (AI=2), showing partial development of stamens and no pollen; and hermaphrodite (AI=3), showing complete development of stamens and pollen. The average AI of each plant and genotype was then assessed from the resulting AI score of at least five individual pistillate flowers from each plant, using a minimum of 30 plants for each genotype.

The growth rates of ovaries and petals in both pistillate and male flowers in each of the WT and mutant plants were assessed by measuring the length of ovaries and petals every 3 days

in at least 12 flowers of each genotype, starting with flowers ~2 mm in length. The anthesis time was estimated as the number of days taken for a 2 mm pistillate or male floral bud to reach anthesis. The effect of the *etr1a* and *etr2b* mutations on the vegetative vigor of each plant was assessed by determining the average plant height, the total number of nodes, and the average internode length in the main shoot of WT and mutant plants at 60 days after transplantation.

4.3.3. Identification of *etr1a* and *etr2b* mutations by whole-genome sequencing analysis

To identify the causal mutations of the *etr1a* and *etr2b* phenotypes, WT and mutant plants, which were both derived from BC_2S_1 segregating populations, were subjected to whole-genome sequencing (WGS). In total, 120 BC_2S_1 seedlings from each mutant family were subjected to the ethylene triple-response assay, and the plants exhibiting the ethylene-insensitive mutant phenotype were separated from the WT. The phenotype of those seedlings was verified in the adult plants, since WT plants were monoecious while homozygous mutant plants were andromonoecious or partially andromonoecious.

The DNA of each plant was isolated by using the Gene JET Genomic DNA Purification Kit (Thermo Fisher), and the DNA from 20 plants showing the same phenotype in each mutant family were pooled into different bulks. Two DNA bulks were generated for each mutant family: one DNA WT bulk and one DNA mutant bulk. The four DNA bulks were used to prepare paired-end multiplex libraries by using the KAPA Library Preparation kit (Kapa Biosystems) and sequenced in 126 base-read pairs on an Illumina HiSeq2000 instrument, following the manufacturer's protocol. Base calling and quality control were done using the Illumina RTA sequence analysis pipeline, according to the manufacturer's instructions. The average fold-effective coverage of all the samples was between 15.5 and 17.2, and the percentage of genomic bases with a fold coverage higher than 5 was between 78.4 and 83.3% (Supplementary Table 4.1). Raw reads were mapped on to the *C. pepo* genome v3.2 by using the GEM mapper (Marco-Sola et al., 2012) in order to generate four bam files. Indels were realigned with GATK, and duplicates were marked with Picard v1.110. In each sample we counted all the 'good-quality bases', selected according to the following filter parameters: base quality ≥17, mapping quality ≥20, read depth ≥8. Variants were called using GATK's HaplotypeCaller in gVCF mode (McKenna et al., 2010), and then the GenotypeGVCF tool was used to perform joint genotyping on all the samples together. Only the single nucleotide

polymorphisms (SNPs) were selected by using GATK's SelectVariants tool. Sites with multiple nucleotide polymorphisms were filtered out.

With regard to the segregation analysis of each mutant phenotype, the genotype of WT plants should be considered as 0/0, as the reference genome. The mutant bulks were completed by using DNA from the most ethylene-insensitive plants. However, given the semi-dominant nature of the *etr1a* and *etr2b* mutations, we expected some plants to be heterozygous for the mutations (0/1), although most of them were expected to be homozygous (1/1), with a mutant allelic frequency (AF) close to 1. Therefore, we filtered our data according to the following parameters: genotype quality ≥20, read depth ≥4 (this specific base was covered with at least four reads), AF ≥0.8 in the two mutant bulks, and AF ≥0.2 in the WT bulks.

4.3.4. Validation of the identified mutations by high-throughput genotyping of individual segregating plants

To confirm that the identified mutations were the causal mutations of the *etr1a* and *etr2b* phenotypes, more than 200 individual BC_2S_1 segregating populations were genotyped by using real-time PCR with TaqMan probes. The multiplex PCRs were done using the Bioline SensiFAST™ Probe No-ROX Kit, a set of forward and reverse primers amplifying the polymorphic sequence, and two allele-specific probes descriptive of the SNP of interest (C-T). The WT probe was labelled with FAM dye, while the mutant probe was labeled with HEX reporter dye. The BHQ1 quencher molecule was used in both probes (Supplementary Table 4.2). After genotyping, plants were phenotyped for ethylene triple response in etiolated seedlings and for sex expression in adult plants, enabling us to see whether the mutant alleles co-segregated with the *etr1a* or the *etr2b* mutant phenotype.

4.3.5. Assessment of relative gene expression by quantitative RT-PCR

Gene expression analysis was performed by using quantitative reverse transcription (qRT)‑PCR on three biological replicates, each one resulting from an independent extraction of total RNA from samples that were pooled from at least three plants with the same genotype.

RNA was isolated according to the protocol of the GeneJET Plant RNA Purification Kit (Thermo Fisher). cDNA was then synthesized by using the RevertAid RT Reverse Transcription Kit (Thermo Fisher). The expression levels of genes were evaluated through qRT‑PCR by using the Rotor-Gene Q thermocycler (Qiagen) and SYBR® Green Master Mix (BioRad). Supplementary Table 4.3 shows the primers used for qRT‑PCR reactions for each analyzed gene.

4.3.6. Bioinformatics and statistical analyses

Alignments were performed using the BLAST alignment tools at NCBI (http://www.blast.ncbi.nlm.nih.gov/) and Clustal Omega at EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/).

The phylogenetic relations of ETR1, ETR2 and ERS1-like ethylene receptors were studied using MEGA7 software (Kumar et al., 2016), which allowed the alignment of proteins and the construction of phylogenetic trees using the Maximum Likelihood method based on the Poisson correction model (Zuckerkandl and Pauling, 1965), with 2000 bootstrap replicates. The protein sequences (Supplementary Table 4.4) were obtained using The Arabidopsis Information Resource (https://www.arabidopsis.org/) and Cucurbit Genomics Database (http://cucurbitgenomics.org/).

Multiple data comparisons were obtained by analysis of variance with the significance level $P<0.05$, and each two averages were compared using Fisher's least significant difference method.

4.4. Results

4.4.1. *etr1a* and *etr2b*, two semi-dominant ethylene-insensitive mutations affecting sex determination

etr1a and *etr2b* are two independent, ethylene-insensitive mutant families that were found after screening a mutant collection of *C. pepo* for ethylene triple response (García et al., 2018). To ensure accurate phenotyping, mutant plants were backcrossed with the background genotype MUC16 for two generations, and then selfed. The resulting BC_1S_1 or BC_2S_1 generations segregated for three ethylene-response phenotypes in etiolated seedlings: ethylene-sensitive plants (WT), intermediate plants (wt/*etr*), and ethylene-insensitive plants (*etr*) (Figure 1).

The segregation ratio of the three ethylene triple response phenotypes in BC_1S_1 and BC_2S_1 generations indicated that the two ethylene-insensitive mutations were semi-dominant, and that the intermediate phenotype corresponded to heterozygous plants (wt/*etr1a* or wt/*etr2b*) (Supplementary Table 4.5) (García et al., 2018). WT plants responded to ethylene with reduced length of the roots and the hypocotyl, and also increased hypocotyl thickness, in comparison to those grown in air. In contrast, the length of the hypocotyl and roots of ethylene-treated homozygous *etr1a* and heterozygous *etr2b* seedlings was reduced to a lesser

extent in relation to those grown in air (Figure 1); this indicated that the level of insensitivity of the two mutants was not total, as they possessed a partially ethylene-insensitive phenotype.

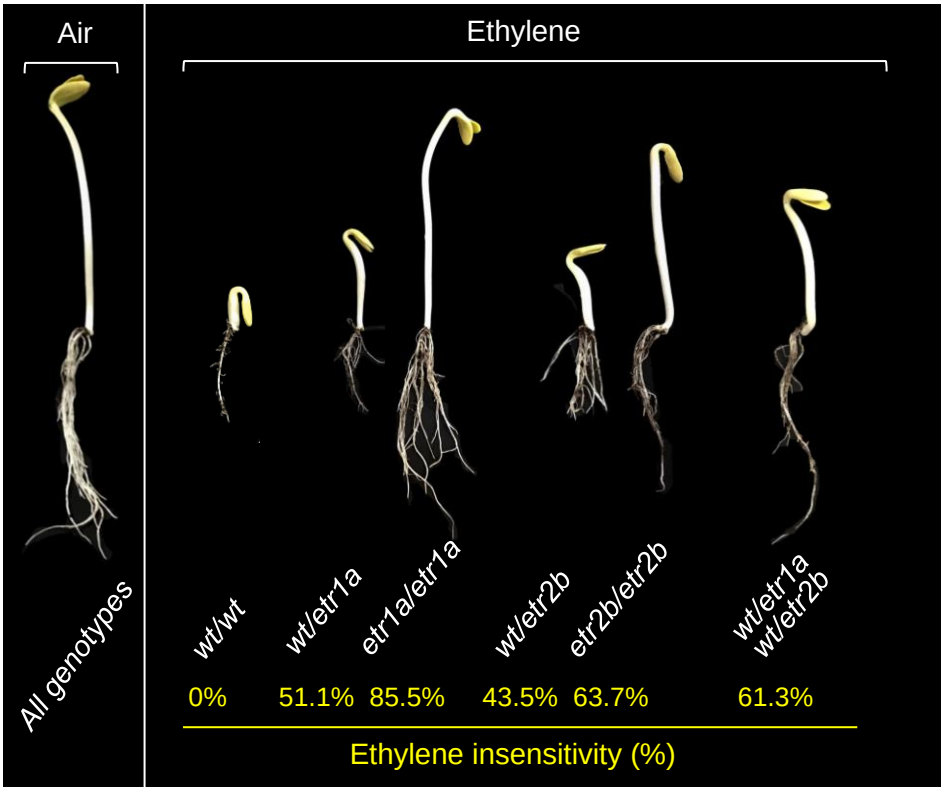


Figure 1. Ethylene triple-response phenotypes of WT and heterozygous and homozygous single and double mutants for *etr1a* and *etr2b*. When grown in air, both the WT and the mutants showed the same growth. When exposed to ethylene, the WT responded with drastic reductions in the length of the hypocotyl and roots, while the mutants had a minor but differential response. The average level of ethylene sensitivity (ES) was assessed as the percentage of reduction in hypocotyl length relative to air-grown seedlings, and assuming that WT is 100% ethylene sensitive. The percentage ethylene insensitivity was then calculated as (100–ES). Assessments were performed in three replicates with at least 20 seedlings for each genotype.

Assuming that WT plants are completely ethylene sensitive (0% ethylene insensitivity), we estimated the percentage of ethylene insensitivity in homozygous and heterozygous seedlings of the two mutants. The homozygous *etr1a* seedlings showed a more severe ethylene-insensitive phenotype than the homozygous *etr2b* seedlings (85.5% ethylene insensitivity for *etr1a/etr1a* versus 63.7% in *etr2b/etr2b*). Heterozygous mutants displayed an intermediate percentage of ethylene insensitivity (51.1% in wt/*etr1a* and 43.5% in wt/*etr2b*) (Figure 1).

Given that ethylene is the main regulator of sex determination in cucurbits, we assessed whether the *etr1a* and *etr2b* mutations altered the sexual phenotype of the plants. Staminate

and pistillate flowers in the first 40 nodes of WT, heterozygous, and homozygous mutant plants were assessed (Supplementary Figure 4.1). All plants showed two sexual phases of development. In the first phase, plants produce only male flowers. The second phase starts after the pistillate flowering transition and is characterized by the production of male and pistillate flowers alternately (Figure 2A). The duration of these two phases of sexual development was affected by the *etr* mutations. Homozygous *etr1a* and *etr2b* mutants showed a significant delay in pistillate flowering transition, as well as a reduction in the number of pistillate flowers per plant (Figure 2B, C). Heterozygous *etr1a* and *etr2b* mutants showed an intermediate phenotype for both traits (Figure 2B, C), confirming the semi-dominant nature of the two mutations.

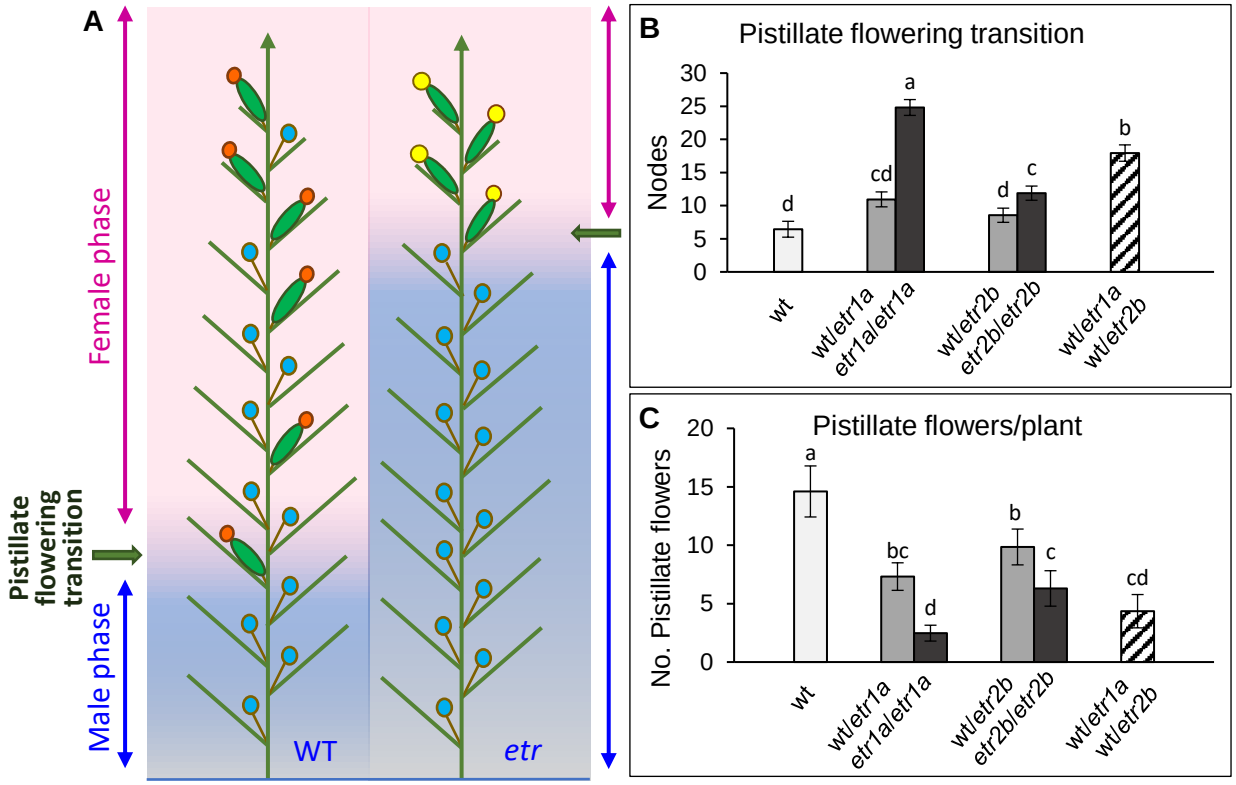


Figure 2. Effect of *etr1a* and *etr2b* on sex expression. (**A**) Schematic representation of the distribution of male and female flowers in WT and *etr1a* or *etr2b* (*etr*) mutant plants. The male and female phases of development, and the node at which plants start to produce pistillate flowers (pistillate flowering transition), are indicated. Blue = male flower; red = female flower; yellow = bisexual or hermaphrodite flower. (**B**) Comparison of pistillate flowering transition in WT plants and single and double mutants. (**C**) Comparison of the number of pistillate flowers per plant in WT and single and double mutant plants. Error bars represent SE. Different letters in (**B**) and (**C**) indicate statistically significant differences ($P<0.05$) between samples.

The most evident phenotypic alteration in the *etr1a* and *etr2b* mutants was the conversion of monoecy into partial or complete andromonoecy. WT plants produced exclusively female flowers after the pistillate flowering transition, indicating a complete arrest of stamen development in pistillate flowers (Figure 3A).

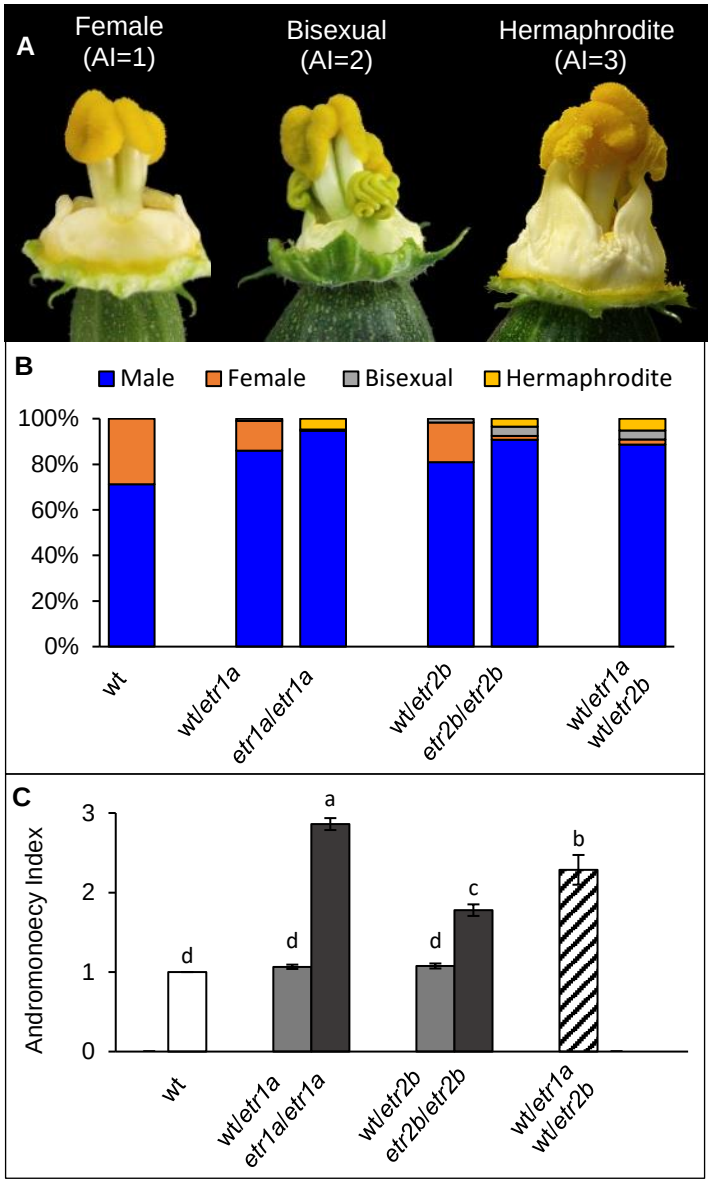


Figure 3. Effect of *etr1a* and *etr2b* mutations on sex determination in *C. pepo*. (**A**) Phenotypes of pistillate flowers in WT and mutant plants: female [andromonoecy index (AI) =1], bisexual (AI=2), and hermaphrodite (AI=3). Female flowers develop no stamen, while bisexual and hermaphrodite flowers develop immature stamens and mature stamens with pollen, respectively. (**B**) Percentage of male, female, bisexual, and hermaphrodite flowers in WT plants and *etr1a* and *etr2b* single and double mutants. (**C**) AI of WT plants and *etr1a* and *etr2b* single and double mutants. The average AI of each genotype was assessed in at least five individual pistillate flowers from each plant, with each genotype containing a minimum of 30 plants. AI varies between 1 (complete monoecy) and 3 (complete andromonoecy). Error bars represent SE. Different letters indicate statistically significant differences (P<0.05) between samples.

In the homozygous *etr1a* mutants, almost all female flowers were transformed into hermaphrodite flowers, showing fully developed stamens with fertile pollen (Figure 3B). In *etr2b* mutants, a high percentage of female flowers were transformed into bisexual (pistillate flowers with immature stamens) or hermaphrodite flowers; however, a small number of female flowers remained (Figure 3B). In heterozygous plants bearing both mutations, only a very small proportion of the female flowers were converted into bisexual flowers (Figure 3B), indicating that with regard to the sex of the flower, *etr1a* and *etr2b* mutations are recessive, and therefore not semi-dominant as shown for ethylene triple response and sex expression.

Given that the degree of stamen development in pistillate flowers was variable, plants were classified according to the AI, ranging from AI=1 (monoecious) to AI=3 (andromonoecious) (Figure 3A). All ethylene-sensitive WT plants showed an average AI of 1. Homozygous plants containing either *etr1a* or *etr2b* mutations displayed an average AI of 2.8 and 1.8, respectively, indicating that the *etr1a* mutation rendered plants almost completely andromonoecious, while the *etr2b* mutation rendered them partially andromonoecious. The average AI for heterozygous plants was 1.1 for both *etr1a* and *etr2b* types, very similar to that of WT plants (Figure 3C).

4.4.2. *etr1a* and *etr2b* mutations alter petal and ovary/fruit development, and affect plant vigor

Table 1 and Figure 4 show the effects of the *etr1a* and *etr2b* mutations on petal and ovary/fruit development. In the hermaphrodite flowers of *etr1a*, the petal growth rate was reduced and resembled petal development in male flowers. Petal maturity and subsequent anthesis of the flower were delayed (Figure 4B). Anthesis time, which is the period of time taken for a 2 mm floral bud to reach anthesis and to open, was longer in male WT flowers (average 22 days) than in female WT flowers (average 14 days) (Table 1). Hermaphrodite *etr1a* flowers took an average of 22 days to reach anthesis (range 15‑40 days). Under the greenhouse conditions used, several hermaphrodite flowers did not reach anthesis; the petals remained green and closed for more than 40 days (Figure 4B). The development to maturity of petals in bisexual flowers of *etr2b* was also delayed in comparison with WT female flowers. However, the delay was less pronounced than in *etr1a* (Figure 4B, Table 1). No alterations in petal development or anthesis time were observed in *etr1a* and *etr2b* male flowers, and no change was found in petal development or anthesis time of female flowers of *etr1a* or *etr2b* heterozygous plants (Figure 4B; Table 1).

Table 1. Anthesis time of pistillate and male flowers in *etr1a* and *etr2b* mutant families.

Family	Flower	Anthesis time (days)		
		wt/wt	***wt/etr***	***etr/etr***
etr1a	Pistillate	14.3±1.2 [b]	14.8±1 [b]	22.7±1.4 [a]
	Male	22.4±0.7 [a]	23.1±1.8 [a]	22.7±1.8 [a]
etr2b	Pistillate	13.8±1.3 [b]	14.3±1.1 [b]	20±4.2 [a]
	Male	22.4±0.5 [a]	21.6±1.1 [a]	22±1.7 [a]

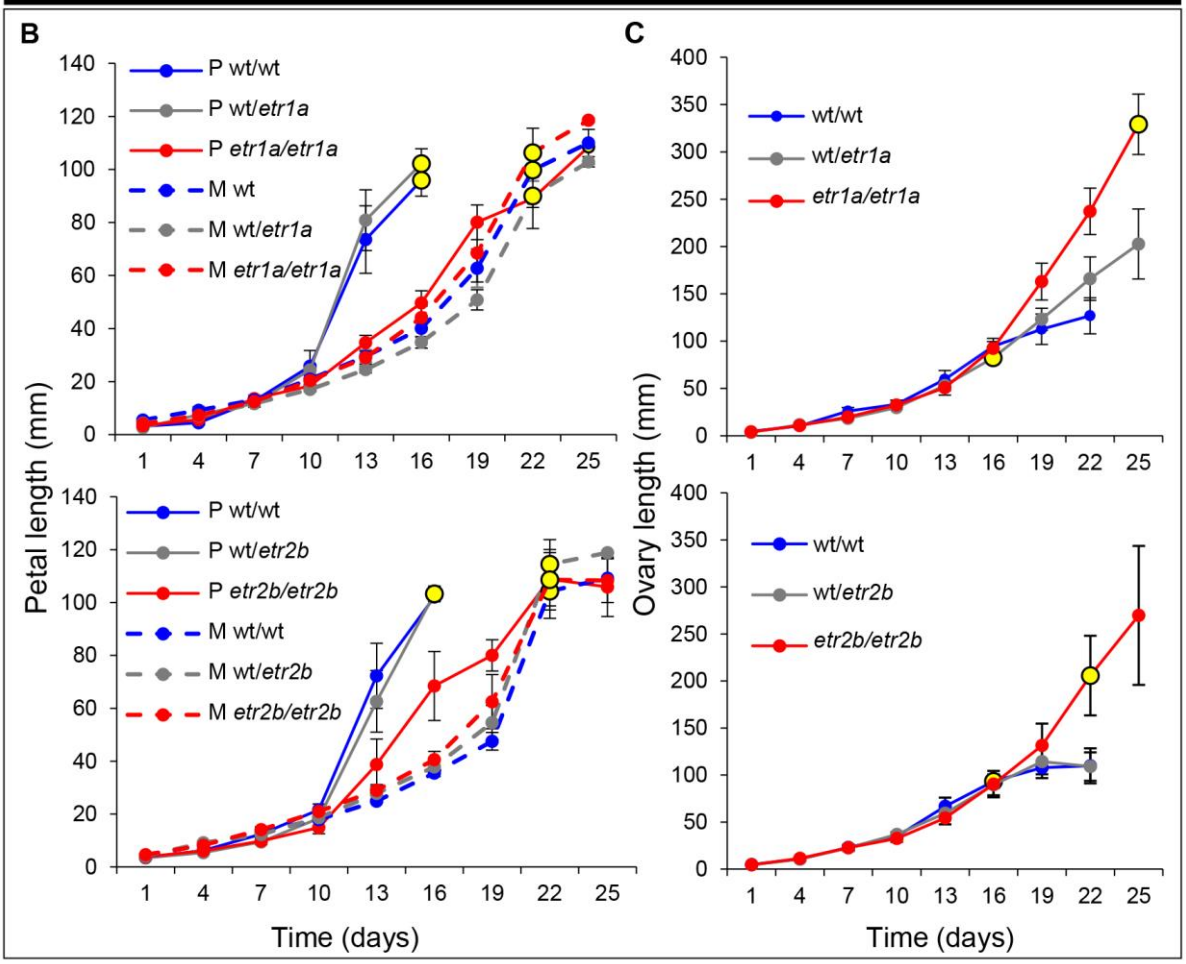
Different letters within the same row indicate significant differences between means (P<0.05).

Significant differences in ovary size were detected between WT and *etr1a* and *etr2b* pistillate flowers (Figure 4). While WT ovaries reached ~8 cm at anthesis and then aborted, the flowers of *etr1a* and *etr2b* reached 16‑40 cm (or even more) at anthesis (Figure 4C). These can be considered as parthenocarpic fruits because the flowers remained closed, rendering them unavailable for pollination. During the first 16 days, the growth rates of WT and mutant ovaries were similar. After 16 days, WT ovaries aborted, while those of homozygous *etr1a* and *etr2b* flowers maintained their growth up to anthesis.

Since the anthesis time of *etr1a* and *etr2b* pistillate flowers was achieved much later than that of WT plants, the size of the mutant ovary at anthesis was much larger (Figure 4C). The ovary growth rate of heterozygous wt/*etr1a* flowers was intermediate between that of wt/wt and homozygous *etr1a/etr1a* (Figure 4C), while heterozygous wt/*etr2b* ovaries displayed the same growth rate as WT plants.

Figure 4. Effect of *etr1a* and *etr2b* mutations on the growth rate of petals and ovaries of male and pistillate flowers. (**A**) Images of pistillate flowers in WT and either *etr1a* or *etr2b* mutants (*etr*) at 7, 14, 20 or 24 days after the flower's ovary reached 4 mm in length. The WT pistillate flower reached anthesis at ~15 days, and in the absence of pollination, the floral organs abscised, and the fruit aborted 2–3 days after anthesis. In the two *etr* mutants, the anthesis time was markedly delayed, but the fruit grew normally in the absence of pollination (parthenocarpic fruits). (**B**) Comparison of the growth rate of WT and mutant petals. Flowers were labeled when their ovaries were 4 mm long, and then measured every 3 days for 25 days. P = pistillate flowers, M = male flowers. Yellow circles indicate the time at which more than 80% of the flowers reached anthesis. (**C**) Comparison of the growth rate of WT and mutant ovaries/fruits over a period of 25 days. Error bars represent SE. (Next page)


A
40 mm
wt
etr
wt
etr
wt
etr
wt
etr
7 days
14 days
20 days
24 days


B
C
P wt/wt
P wt/etr1a
P etr1a/etr1a
M wt
M wt/etr1a
M etr1a/etr1a
P wt/wt
P wt/etr2b
P etr2b/etr2b
M wt/wt
M wt/etr2b
M etr2b/etr2b
Petal length (mm)
Ovary length (mm)
wt/wt
wt/etr1a
etr1a/etr1a
wt/wt
wt/etr2b
etr2b/etr2b
Time (days)
Time (days)

Pollination was attempted in homozygous mutant flowers that reached anthesis, but none of the fruits in the homozygous *etr1a* and *etr2b* mutant plants were able to set seeds under the given conditions. Since the pollen of these mutants is fertile, these results indicate female sterility associated with the *etr1a* and *etr2b* mutations. As no pure seed could be obtained from *etr1a* or *etr2b* homozygous mutants, they were maintained by selfing heterozygous mutant plants.

Vegetative development was enhanced in mutant adult *etr1a/etr1a* and *etr2b/etr2b* plants. Table 2 shows the differences in plant height, number of nodes, and internode length of plants grown under the same conditions. The two mutations were associated with increased plant growth rate and height compared with WT plants. These differences in height and growth rate were mainly due to an increase in the internode length, which was approximately twice as high in mutants as in WT plants. The number of nodes developed by WT and mutant plants was, however, very similar (Table 2).

Table 2. Effects of the *etr1a* and *etr2b* mutations on plant vegetative development.

	Plant height (cm)	**Node number**	**Internode length**
WT	85.43 [b]	47.2 [b]	13.5 [b]
etr1a/etr1a	112.5 [a]	50.7 [a]	24.1 [a]
WT	93.9 [b]	47.1 [a]	14.8 [b]
etr2b/etr2b	113.3 [a]	48.5 [a]	28.8 [a]

Different letters indicate significant differences between wild-type (WT) and homozygous mutants ($P<0.05$).

4.4.3. Phenotype of *etr1a* and *etr2b* double mutants

The interaction between the *etr1a* and *etr2b* mutations was studied in double-heterozygous plants for both mutations. As mutant bisexual and hermaphrodite flowers were sterile, heterozygous wt/*etr1a* plants were pollinated with pollen from homozygous *etr2b*/*etr2b* plants. The progeny segregated as 1:1 for single-heterozygous wt/wt wt/*etr2b* and double-heterozygous wt/*etr1a* wt/*etr2b* plants. However, double-heterozygous plants were also female sterile, which made it impossible to generate double-homozygous plants bearing the two mutations. The reciprocal cross produced the same results.

Double-heterozygous mutants showed a more severe ethylene-insensitive phenotype (61.3% insensitivity) than single-heterozygous mutants (51.1% insensitivity for wt/*etr1a* and 43.5%

for wt/*etr2b*), and their phenotype resembled the triple response of *etr2b/etr2b* (Figure 1). The maleness effect of the *etr1a* and *etr2b* mutations was also enhanced in the double-heterozygous wt/*etr1a* wt/*etr2b* plants compared with the phenotype of the single-heterozygous mutants. The female flowering transition was delayed up to an average of 17.9 nodes, and pistillate flower production was also significantly reduced to about 4.3 flowers resulting from the first 40 nodes of the plant, again resembling the phenotype exhibited by the homozygous single mutants (Figure 2B, C).

In the double-heterozygous wt/*etr1a* wt/*etr2b* plants, most female flowers were converted into bisexual or hermaphrodite flowers (average AI=2.3), contrasting with single-heterozygous plants, which produced nearly 100% female flowers (average AI=1.1) (Figure 3B, C). Thus, the combination of the two *etr* mutations, albeit in heterozygous conditions, had a similar effect on the flower sexual phenotype as the homozygous single mutations (Figure 3B, C). These data therefore indicate that the two mutations have an additive effect on the sex phenotype of pistillate flowers, and that the effect is dependent upon the number of mutant alleles for the two loci.

4.4.4. Identification of the *etr1a* and *etr2b* mutations

The causal mutations of the *etr1a* and *etr2b* phenotypes were identified by WGS. One DNA WT bulk and one DNA mutant bulk were made for each mutant family, each bulk consisting of a pool of DNA from 20 plants from the same BC_2S_1 segregating population. With regard to the mutant bulk, only the ethylene-insensitive plants that showed the most strongly andromonoecious phenotype were selected, in order to assure that they were homozygous for the mutations. The four DNA bulks were subjected to WGS and the results were mapped against the *C. pepo* reference genome version 3.2. The identified SNPs were filtered by using different criteria. Common variants between samples of the two families were discarded, as they were considered likely to correspond to spontaneous nucleotide polymorphisms in the MUC16 genetic background (García et al., 2018). SNPs corresponding to canonical EMS mutations (C>T and G>A) were selected, and filtered for their quality and depth, as well as for their mutant allele frequency (AF), in WT and mutant DNA samples. We expected the WT bulks to present the genotype 0/0 (AF=0) and the mutant bulks to present the genotype 1/1 (AF=1).

Assuming minor contamination of the bulks with some heterozygous plants, we discarded SNPs with AF<0.8 in the mutant samples and AF>0.2 in the WT samples. After filtering, two EMS candidate mutations were selected for *etr1a* and one candidate mutation was

selected for *etr2b* (Supplementary Table 4.1). The sequences surrounding the candidate mutations (±500 bp) were then used in BLAST searches against the DNA and protein databases at NCBI, which detected an EMS canonical C>T transition in both *etr1a* and *etr2b* families, located in the coding region of the ethylene receptor genes *CpETR1A* and *CpETR2B*, respectively (Table 3).

Table 3. Genotype, depth (DP), and allele frequency (AF) of causal mutations of the *etr1a* and *etr2b* phenotypes.

	DNA bulk	**Causal mutations**					
		LG07: 6 891 436 (C>T)			**LG03: 11 824 350 (C>T)**		
		Genotype	**DP**	**AF**	**Genotype**	**DP**	**AF**
etr1a	WT	0/0	20	0.15	0/0	19	0
	Mutant	1/1	20	1	0/0	9	0
etr2b	WT	0/0	16	0	0/0	10	0
	Mutant	0/0	8	0	1/1	14	1

To verify that these were the causal mutations of the *etr1a* and *etr2b* phenotypes, more than 200 plants segregating for either *etr1a* or *etr2b* were genotyped for the WT and mutant alleles of *CpETR1A* and *CpETR2B*. The results demonstrated a perfect co-segregation of the *etr1a* and *etr2b* sex phenotypes with mutations in *CpETR1A* and *CpETR2B*, respectively (Table 4). The other candidate mutation for *etr1a* segregated independently of the mutant phenotype.

With regard to the ethylene triple response, the plants that were homozygous for the WT alleles were sensitive to ethylene, and those that were homozygous for the mutant alleles of *CpETR1A* or *CpETR2B* were ethylene insensitive; heterozygous plants showed intermediate triple-response phenotypes (Table 4). Moreover, plants that were homozygous for any of the identified mutations were all andromonoecious, while those that were either homozygous for the WT allele or heterozygous were all monoecious (Table 4). This finding demonstrates that mutations identified in the two ethylene receptor genes *CpETR1A* and *CpETR2B* co-segregated with the ethylene-insensitive phenotype and andromonoecy of the *etr1a* and *etr2b* mutants.

Table 4. Co-segregation analysis of the *CpETR1A* and *CpETR2B* mutations with the ethylene-insensitive phenotypes in BC_2S_1 populations segregating for *etr1a*, *etr2b*, and double mutants

Segregating population	*CpETR1A* or *CpETR2B* genotypes	Triple response to ethylene			Sexual phenotype		Total
		Sensitive	Intermediate	Insensitive	Monoecious	Andromoecious	
etr1a	0/0	80	-	-	80		**226**
	0/1	-	61		61		
	1/1	-		85		85	
etr2b	0/0	85	-	-	85		**234**
	0/1	-	62		62		
	1/1	-		87		87	
Double mutants	0/0; 0/1	-	22		22		**47**
	0/1; 0/1	-		25		25	

4.4.5. Gene structure of *CpETR1* and *CpETR2*

The *de novo* assembly of the *C. pepo* genome, published recently, revealed a whole-genome duplication, which occurred just before the speciation event that created the genus *Cucurbita* (Montero-Pau et al., 2017; Sun et al., 2017). Accordingly, we found that the genomes of the *Cucurbita* species contain two paralogs for each of the ethylene receptors ETR1, ERS1, and ETR2. No ERS2- or EIN4-like receptors were found in the genomes of these species. The *C. pepo* genome has two ETR1 duplicates (*CpETR1A* and *CpETR1B*), which showed more than 90% homology and mapped on chromosomes 7 and 11, and two ETR2 duplicates (*CpETR2A* and *CpETR2B*), which showed more than 89% homology and mapped on chromosome 8. The duplicates maintained the same molecular structure: six exons and five introns for the two *CpETR1* paralogs, and three exons and two introns for the two *CpETR2* paralogs (Figure 5A, B).

The *etr1a* mutation was located at nucleotide position 284 of the first exon of the *CpETR1A* gene, and the *etr2b* mutation was located at nucleotide position 1018 of the first exon of the *CpETR2B* gene (Figure 5A, B). The deduced ETR1A and ETR2B receptors had the same domains as those of the Arabidopsis homologs: an ethylene-binding domain with three transmembrane segments; one GAF domain; one histidine-kinase domain; and a response regulator receiver domain. The two mutations resulted in an amino acid substitution in each protein: A95V in the ethylene-binding domain of CpETR1A, and E340K in the coiled-coil domain between the GAF and histidine-kinase domains of CpETR2B (Figure 5A, B). All

ETR1- and ETR2-like proteins in the NCBI database were found to contain the WT amino acid at these two particular positions, indicating that the amino acids affected by the two mutations are highly conserved in very different plant species (Supplementary Figures 4.2 and 4.3).

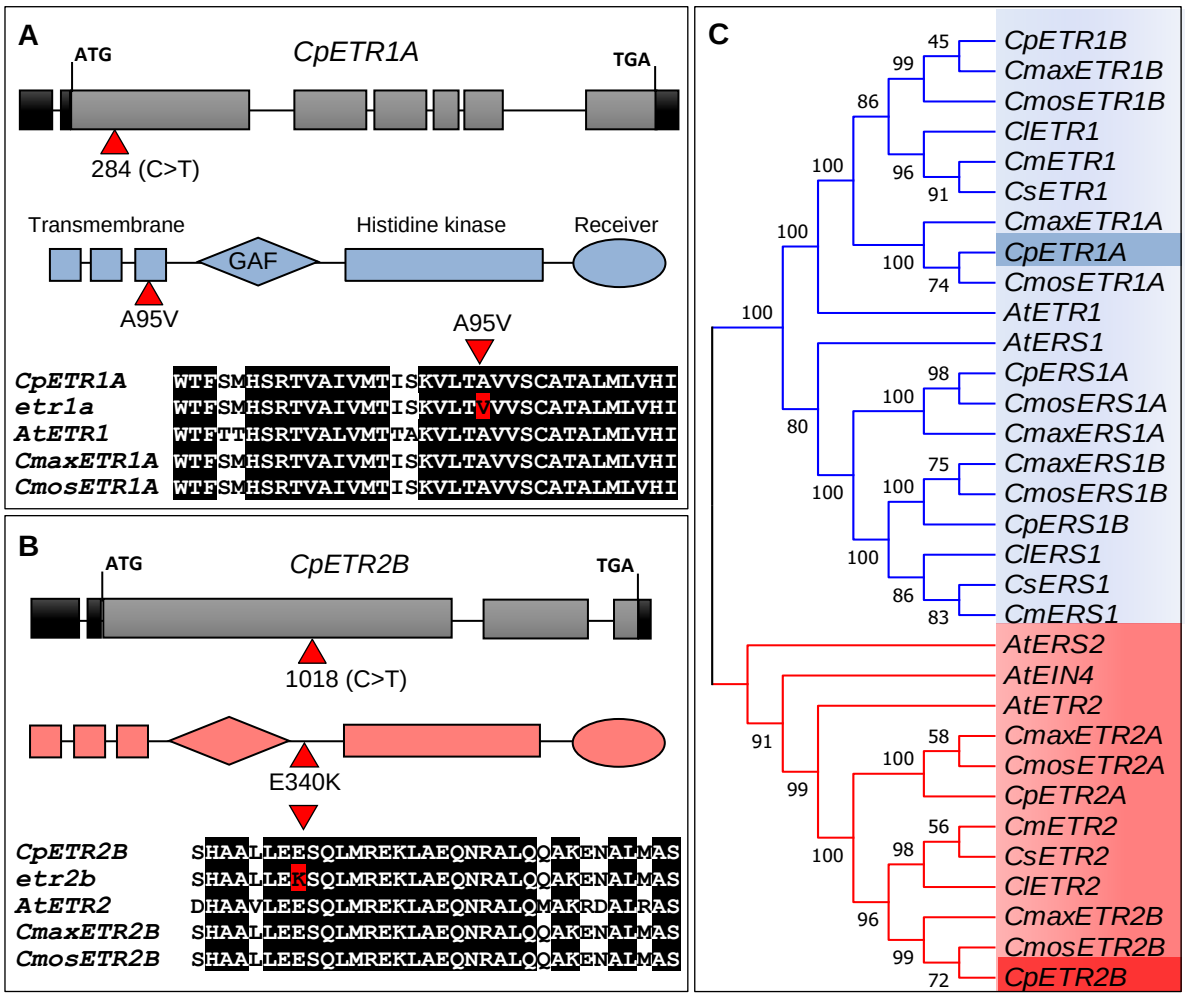


Figure 5. Molecular structure of CpETR1A and CpETR2B, and phylogenetic analysis of ETR1- and ETR2-like ethylene receptors. (**A**) *CpETR1A*; (**B**) *CpETR2B*. Within each gene, black boxes indicate untranslated regions and grey boxes correspond to exons. Numbers indicate the size of the fragment (bp). Missense mutations in the gene, and amino acid substitutions in the protein, are indicated in red. EMS mutations C>T are numbered with respect to the coding sequence. The three transmembrane subdomains in the ethylene-binding domain, as well as the GAF, histidine-kinase, and receiver domains, are indicated for each ethylene receptor. (**C**) Phylogenetic relationships between the ethylene receptors of six cucurbit species (*Cucurbita pepo*, *Cucurbita moschata*, *Cucurbita maxima*, *Cucumis melo, Cucumis sativus*, and *Citrullus lanatus*) and those of Arabidopsis. The six ethylene receptors of *C. pepo* are positioned in three clusters corresponding to ETR1, ERS1, and ETR2 of Arabidopsis. In the three *Cucurbita* species, ETR1 and ETR2 are duplicated in relation to Arabidopsis.

To assess the genetic relationships between ETR1- and ETR2-like ethylene receptors in *C. pepo*, a phylogenetic tree was inferred from the ETR1, ERS1, and ETR2 protein sequences of different cucurbit species -*Cucurbita maxima*, *Cucurbita moschata*, *C. melo*, *C. sativus*, and *C. lanatus*- together with the five ethylene receptors of Arabidopsis (Figure 5C, Supplementary Table 4.4). In the other cucurbits, only one single gene was found for each ethylene receptor. In agreement with the allotetraploid origin of the genus *Cucurbita* (Sun et al., 2017), only one of the paralogs in subgenome B (*CpETR1B*, *CpERS1B*, and *CpETR2B* in *C. pepo*) clustered with their orthologs in the rest of the cucurbit species (Figure 5C).

4.4.6. Effects of *etr1a* and *etr2b* on ethylene receptor gene expression

The expression patterns of the mutated *CpETR1A and CpETR2B* and their duplicated paralogs (*CpETR1B* and *CpETR2A*) were studied in leaves, roots, shoots, and shoot apices of WT plants (Figure 6). The four genes were expressed in all the analyzed tissues, suggesting that both duplicates maintain their expression in these tissues.

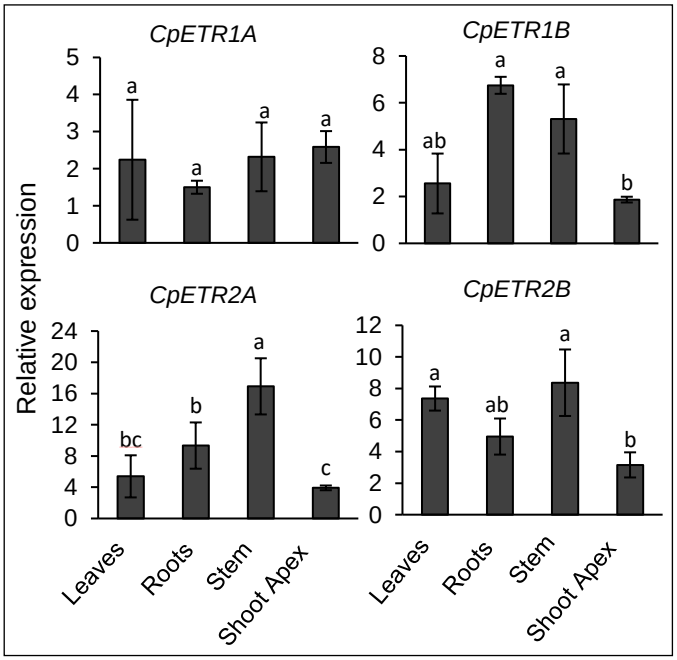


Figure 6. Expression patterns of *CpETR1* and *CpETR2* ethylene receptor genes in different plant organs of *C. pepo* WT plants. The relative level of each transcript was quantified by qRT–PCR in three independent replicates of each tissue. The shoot apex consists not solely of the apical meristem but also of small leaves and floral buds. Different letters indicate statistically significant differences ($P<0.05$) between samples.

We also investigated whether the *etr1a* or *etr2b* mutations alter the patterns of expression of the four *ETR* genes. Gene expression was compared in WT and mutant female floral buds at

very early stages of development, that is, when stamen arrest takes place in the WT female flower (stage T0), and at 1 day before anthesis of WT flowers (stage T5); in the latter case, we separated the flower into the ovary and a tissue comprising the petals, style, and stigma (Figure 7). The *etr1a* mutation inhibited the expression of *CpETR1A* and *CpETR2A* in the tissue that comprised the petals, style, and stigma of T5 flowers, and also reduced the transcription of *CpETR2B* in T0 flowers (Figure 7). By contrast, the *etr2b* mutation inhibited the expression of *CpETR1B* and *CpETR2A* in T0 floral buds, and the expression of *CpETR1A* in the ovary of T5 flowers.

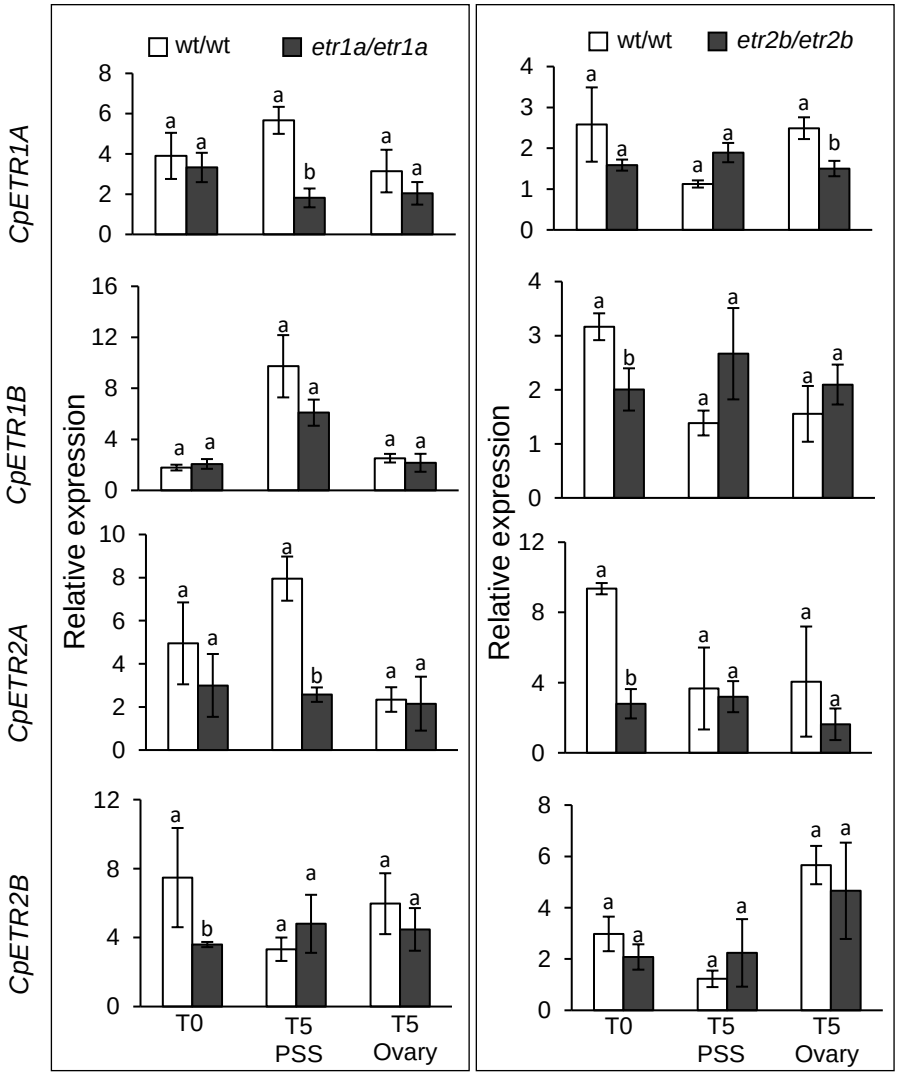


Figure 7. Relative expression of *CpETR1* and *CpETR2* ethylene receptor genes in female flowers of *C. pepo*. The relative level of each transcript was quantified by qRT–PCR in three independent replicates of each tissue. T0 corresponds to complete female flowers 2 mm in length; T5 corresponds to pre-anthesis-stage female flowers, separated into the ovary and a tissue comprising the petals, style, and stigma (PSS). The comparison of gene expression was performed between homozygous WT and homozygous mutant flowers derived from plants in the same segregating population. Different letters indicate statistically significant differences (P<0.05) between samples.

4.5. Discussion

4.5.1. *etr1a* and *etr2b* are two missense mutations in ethylene receptors leading to semi-dominant ethylene insensitivity

Ethylene is perceived by a family of two-component histidine kinase receptors that repress the ethylene signaling cascade in the absence of ethylene but become inactivated upon ethylene binding (Hua and Meyerowitz, 1998). Mutations in ethylene receptor genes fall into two main categories: (i) dominant gain-of-function mutations, conferring ethylene insensitivity, and (ii) recessive loss-of-function mutations that have little effect as single mutations but show a constitutive ethylene response in combination, for example, in double, triple, and quadruple mutants of Arabidopsis. Both *etr1a* and *etr2b* of *C. pepo* correspond to the first type of mutation, since they are semi-dominant and result in plant ethylene insensitivity (Figure 1).

In Arabidopsis dominant mutants, the ethylene-insensitive phenotypes are caused by single amino acid substitutions in the transmembrane ethylene-binding domain of any of the five ethylene receptors described in this species (Bleecker *et al*., 1988; Chang *et al*., 1993; Guzmán and Ecker, 1990; Hua *et al*., 1995, 1998; Wang *et al*., 2006). The *etr1a* mutation described here is also a missense mutation (A95V), situated in the third transmembrane domain of the N-terminal ethylene-binding site of *CpETR1A* (Figure 5), which causes a strong reduction in ethylene sensitivity in etiolated seedlings (Figure 1). The mutation is contained in a conserved segment of the protein, close to the T94M mutation of Arabidopsis ETR1, which is known to disrupt the ability of the receptor to bind ethylene and to strongly affect ethylene sensitivity (Resnick et al., 2008; Wang et al., 2006). The *etr2b* mutation, however, is a missense mutation (E340K) within the coiled-coil domain between the GAF and histidine-kinase domains of *CpETR2B* (Figure 5). Given that this domain does not participate in ethylene binding, it is likely that the ethylene insensitivity of *etr2b* is caused by a lack of transduction of the ethylene signal, as has been suggested for other dominant ethylene-insensitive mutations (Hall et al., 1999). Thus, *etr1a* may disrupt the ethylene-binding site, and *etr2b* may alter ethylene signal transduction, but both mutations should convert the CpETR1A or CpETR2B receptors to a constitutive signaling-on state that represses the ethylene response (Figure 8). The differing levels of ethylene insensitivity shown by single *etr1a* and *etr2b* mutants of *C. pepo* could indicate that *CpETR1A* (subfamily I) has a more prominent role in ethylene perception than *CpETR2B* (subfamily II). In

Arabidopsis, both *etr1-1* and *etr1-4* mutants are insensitive to ethylene, but *etr1-2*, *etr2-1*, and *ein4-3* maintain a reduced response to ethylene (Hall et al., 1999).

A whole-genome duplication occurred just before speciation of the genus *Cucurbita,* which explains why only the species of the genus *Cucurbita* have two subgenomes (A and B) (Montero-Pau et al., 2017; Sun et al., 2017). In *C. pepo*, we found that the paralogs of *CpETR1* (*CpETR1A* and *CpETR1B*) and *CpETR2* (*CpETR2A* and *CpETR2B*) in each subgenome are both expressed and could be functional (Figures 6 and 7). This tetraploid origin of *C. pepo* could account for the semi-dominant nature of *etr1a* and *etr2*b. In Arabidopsis, ethylene-insensitive mutants show the same phenotype in both homozygous and heterozygous conditions; however, an extra WT allele of the same gene in triploid plants (*mut*/wt/wt) for either *etr1-1*, *etr1-2*, or *ein4-3* reduced ethylene insensitivity compared with homozygous or heterozygous diploid mutants (*mut/mut* or wt/*mut)*. This suggests that the dominant insensitivity of these mutant alleles is mediated by interaction between WT and mutated ethylene receptor isoforms (Hall et al., 1999). Our findings also indicate that the CpETR1A and CpETR2B ethylene receptors exert their action cooperatively rather than independently. Different members of the Arabidopsis ethylene receptors form homomeric and heteromeric complexes that may facilitate receptor signal output (Binder and Bleecker, 2003; Gao et al., 2008; Gao and Schaller, 2009; Grefen et al., 2008; Xie et al., 2006), and mutant receptors can repress the ethylene response only when coupled with a WT receptor, which implies that each tissue can have mixed receptor complexes with different receptor signal output strengths (Li et al., 2009b).

The cooperation between WT and mutant alleles of *CpETR1A* and *CpETR2B* is also suggested by the phenotypes of single and double mutants, where homozygous single mutants for *etr1a* or *etr2b* (85% and 63.7% ethylene insensitivity, respectively) show similar ethylene insensitivity to the heterozygous double mutant wt/*etr1* wt/*etr2b* (61.3%) but stronger insensitivity than that of the heterozygous single mutants wt/*etr1a* and wt/*etr2b* (53.5% and 43.5%, respectively). The suppression of the ethylene response in *C. pepo* is therefore dependent on the dosage of WT and mutant alleles for either *CpETR1A* or *CpETR2B* (Figures 8 and 9).

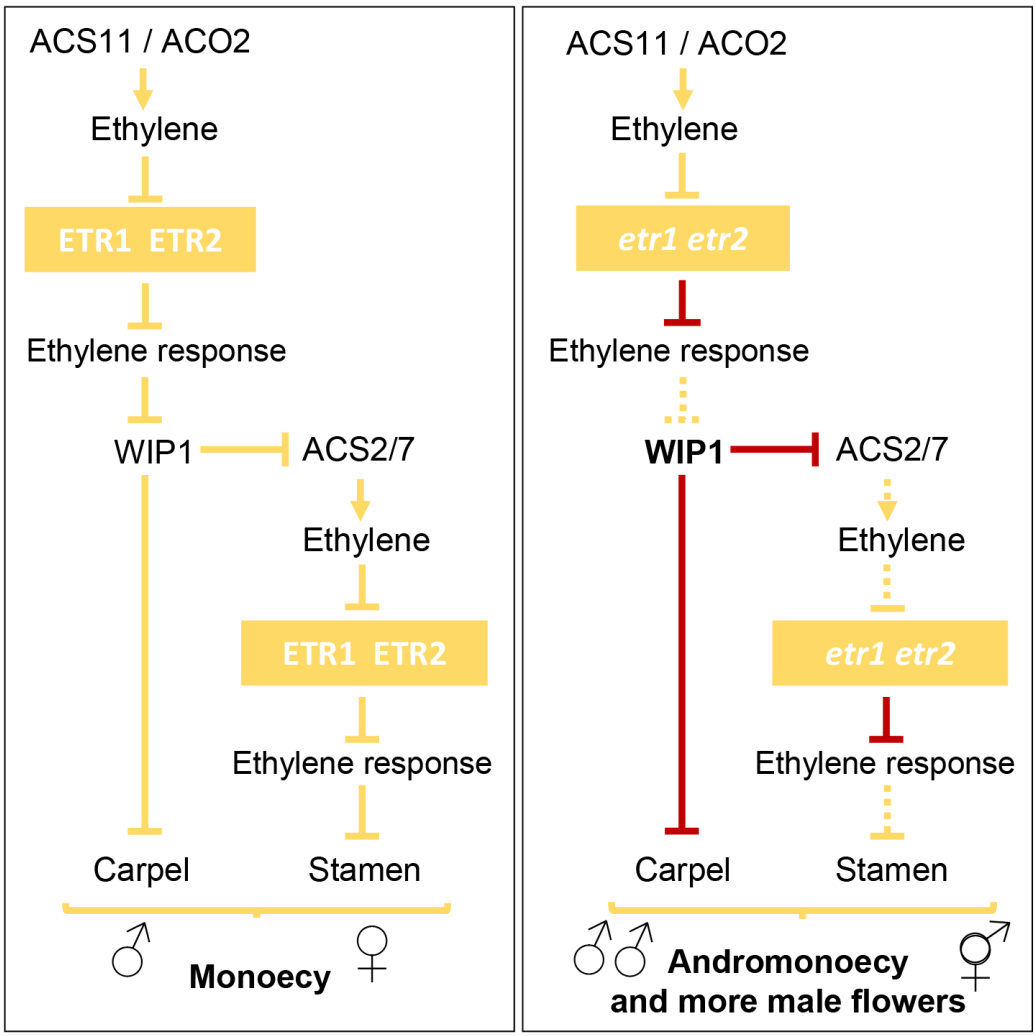


Figure 8. Model of sex determination in cucurbit species, integrating the function of the ethylene receptors ETR1 and ETR2 of *C. pepo* with other sex-determining genes identified in *C. melo* and *C. sativus* (*ACS11, ACO2, WIP1*, and *ACS2/7*). The ethylene biosynthesis enzymes ACS11 and ACS2/7 have different spatiotemporal expression patterns, causing the arrest of carpels or stamens required for a flower to develop as male or female, respectively (left panel). The two biosynthetic pathways are connected by the transcription factor WIP1, which represses the transcription of ACS2/7, and is negatively regulated by the ethylene-producing enzymes ACS11and ACO2 (Boualem et al., 2015; Che and Zhang, 2019). The ethylene receptors ETR1 and ETR2 should perceive and transmit signaling of the ethylene synthesized by both the ACS11 and ACS2/7 pathways. The inhibition of the ACS2/7 ethylene response releases the arrest of stamens in female flowers, resulting in the production of bisexual or hermaphrodite flowers (andromonoecy). On the other hand, the inhibition of the ACS11 ethylene response can induce WIP1, which enhances the arrest of carpels and so leads to the formation of male flowers. Red and dotted lines in the right panel indicate increased or decreased effects, respectively, produced by the *etr1* and *etr2* mutations.

Ethylene insensitivity in the two *etr* mutants was not, however, associated with the expression level of the genes. The *etr1a* and *etr2b* mutations scarcely affected the expression

of ethylene receptor genes, although some of them were downregulated in female flowers, probably owing to a mechanism that aids the recovery of ethylene sensitivity that may be lost as a result of the mutations. Functional compensation between ethylene receptor gene families has been found to occur between *NR* and *LeETR4* of tomato (Tieman et al., 2000) and in certain loss-of-function ethylene receptor mutants of Arabidopsis (Chen et al., 2007; Harkey et al., 2018; Zhao et al., 2002), but not in rice (Wuriyanghan et al., 2009). Nevertheless, given that some of the *ETR* genes were not downregulated in all *etr1a* or *etr2b* tissues, it is also likely that reduction in gene expression reflects a differential regulation of the analyzed ethylene receptor genes.

4.5.2. Mutations in *CpETR1A* and *CpETR2B* alter sex determination and expression

Sex determination in individual floral buds of monoecious and dioecious species is controlled by diverse mechanisms that suppress the development of either stamen or carpel primordia in floral buds that will result in female or male flowers, respectively (Figure 8; Bai et al., 2004; Kater et al., 2001). In the Cucurbitaceae, the main sex regulator is ethylene, but the genetic network controlling sex determination in these species is still poorly understood. With the exception of *WIP1*, all major genes controlling sex determination encode key enzymes involved in ethylene biosynthesis (ACS and ACO; Figure 8). The gene *ACS2/ACS7* controls monoecy, and loss-of-function mutations lead to plants with male and hermaphrodite flowers (andromonoecy) (Boualem et al., 2008, 2009; Ji et al., 2016; Manzano et al., 2016; Martínez et al., 2014). Mutations in this gene do not affect sexual expression, that is, the ratio between male and female flowers in the plant (Manzano et al., 2016; Martínez et al., 2014). On other hand, the genes *ACS11* and *ACO2* are required for the female flower development pathway, and mutations in these genes result in plants with only male flowers (androecy) (Boualem et al., 2015; Chen et al., 2016). Finally, the transcription factor WIP1 is required for male flower development and is negatively regulated by ACS11, and its dysfunction results in plants with only female flowers (gynoecy) (Boualem et al., 2015; Hu et al., 2017).

So far, no ethylene receptor has been positioned in the genetic network controlling sex determination in this group of species (Figure 8), although there was some evidence indicating their participation. Thus, the transcription level of *ETR2* and *ERS1* is higher in gynoecious than monoecious apical shoots of *C. sativus* (Yamasaki et al., 2000), and the down-regulation of *CsETR1* in the stamens of female cucumber flowers appears to be required for the arrest of stamen development (Wang et al., 2010a). Moreover, transgenic *C.*

melo plants overexpressing the Arabidopsis ethylene-insensitive allele *etr1-1* are altered in sex determination and sex expression (Little et al., 2007; Switzenberg et al., 2015). Given that *etr1a* and *etr2b* not only disrupt female flower development (converting monoecy into andromonoecy) but also significantly increase the number of male flowers in the plant, it is likely that ETR1 and ETR2 integrate the two ethylene biosynthesis pathways that result in the determination of male and female flowers, perceiving and signaling the ethylene produced by ACS2/7 as well as that produced by ACS11 and ACO2 (Figure 8).

The mechanisms triggering the sex of each specific floral meristem must be regulated by the level of ethylene sensitivity conferred by ethylene receptors (Figure 9). In fact, the degree of conversion of female flowers into bisexual and hermaphrodite flowers, the delay in pistillate flowering transition, and the increase in the number of male flowers per plant are all correlated with the level of ethylene sensitivity in *etr1a* and *etr2b* single and double mutants.

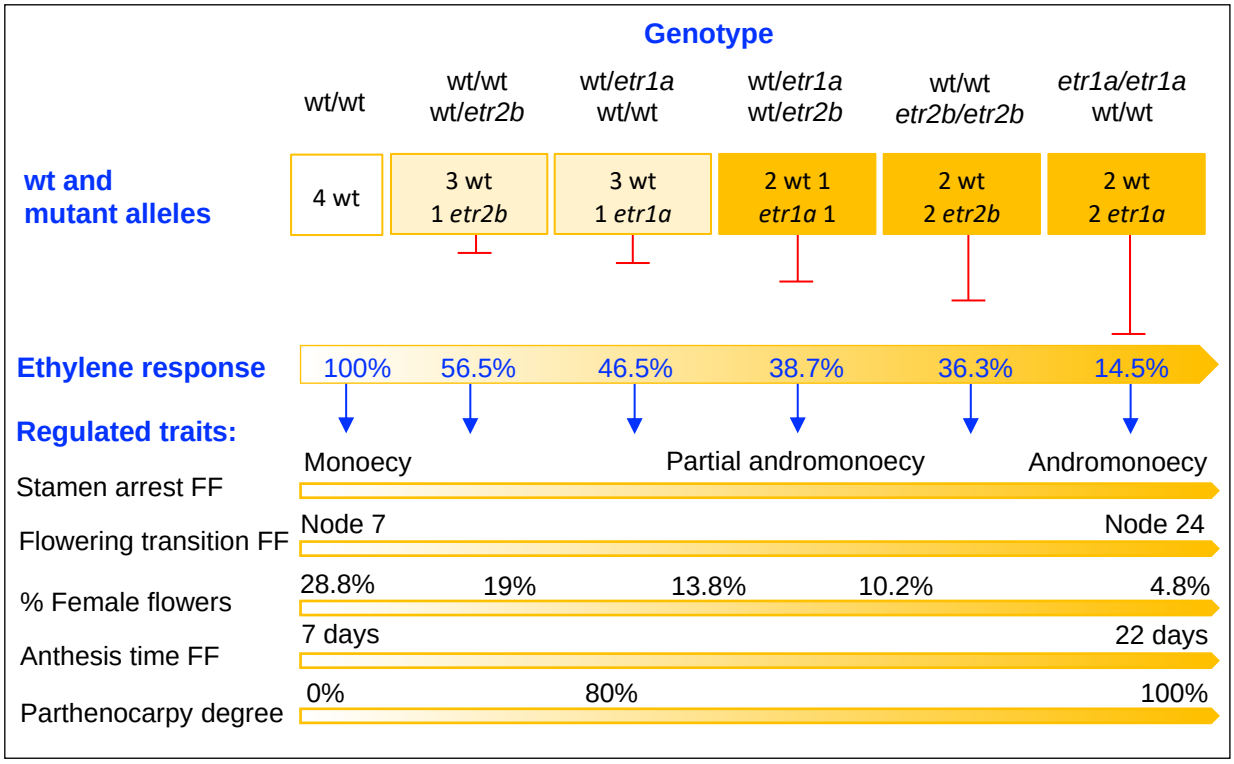


Figure 9. Cooperation of the CpETR1A and CpETR2B ethylene receptors in the regulation of the ethylene response and ethylene-associated traits in *C. pepo* flower and fruit development. In the absence of ethylene, CpETR1A and CpETR2B repress the ethylene response. In the presence of ethylene, the repression of the ethylene response is determined by the dosage of WT and mutant alleles, and also by the strength of each mutant allele, with *etr1a* having a stronger effect than *etr2b*. The resulting ethylene sensitivity regulates the development of pistillate floral organs and sex determination. As the number of mutant alleles increases, the response to ethylene is reduced. This increases the degree of andromonoecy, delays the pistillate (female) floral transition, and decreases the number of pistillate flowers per plant. It also delays the maturity of pistillate flowers, favoring growth of the ovary and parthenocarpic development of the fruit. FF, female flowers.

The final level of ethylene sensitivity in the tissue will be affected by the strength of each mutation, but also by the number of WT and mutant ethylene receptors in the tissue (Hall et al., 1999), and by the cooperation between them for repressing ethylene signaling (Figure 9).

The ethylene-insensitive mutations also alter the growth rate of petals and carpels, making mutant bisexual and hermaphrodite flowers reach anthesis later than WT female flowers. This delay in anthesis time could be associated with the development of stamens, since the anthesis time of the mutant hermaphrodite flower was similar to that of the male flower. Moreover, the mutant ovary continues to grow as long as the petals remain green and do not reach anthesis (Figure 4). These data suggest that fruit set in *C. pepo* is not triggered by pollination, but is a default developmental program, which has a checking point at anthesis. If the flower is pollinated, development of the ovary continues and the fruit sets. In the absence of pollination and fertilization, growth of the ovary is aborted. Fruit set is known to be regulated positively by hormones such as auxins and gibberellins, and negatively by ethylene (Martínez et al., 2013; Shinozaki et al., 2018; Shnaider et al., 2018). Therefore, the reduction of ethylene sensitivity in the *etr1a* and *etr2b* mutants could trigger parthenocarpy by delaying flower anthesis and consequently the checking point for fruit set.

5. Two androecious mutations reveal the crucial role of ethylene receptors in the initiation of the female flower development in *Cucurbita pepo*

5.1. Abstract

Ethylene is the key regulator of sex determination in the monoecious species of the family *Cucurbitaceae*. This hormone determines which individual floral meristem develops as a female or a male flower, and female flowering transition. The sex determination genes discovered up to date code for ethylene biosynthesis enzymes, but little is known about the importance of ethylene signaling components. In this paper we characterize two novel ethylene insensitive mutations (*etr1a-1* and *etr1b*) which block the female flowering transition of *C. pepo;* this makes plants to produce male flowers indefinitely (androecy). Two missense mutations in the ethylene binding domain of the ethylene receptors CpETR1A or CpETR1B were identified as the causal mutations of these phenotypes by using whole-genome re-sequencing. The distinctive phenotypes of single and double mutants for four *etr* mutations, have demonstrated that the final level of ethylene insensitivity depends upon the strength and dosage of mutant alleles for at least three cooperating *ETR* genes, and that the level of ethylene insensitivity determines the final sex phenotype of the plant. Sex phenotype ranges from monoecy in ethylene-sensitive WT plants to androecy in the strongest ethylene-insensitive ones, via andromonoecy in the partially ethylene-insensitive plants. In the stronger mutants, the blocking of female flowering transition was associated with a lack of transcriptional activation of *CpACS11*, *CpACO2* and *CpACS2*/7, three ethylene biosynthesis genes required for female flower development. A model is proposed herewith, integrating both ethylene biosynthesis and receptor genes into the genetic network which regulates sex determination in *C. pepo*.

Key words: squash, ethylene-insensitive mutations, sex determination, monoecy, andromonoecy, androecy, ethylene receptors.

5.2. Introduction

The *Cucurbitaceae* family includes several economically important crop species such as cucumber (*C. sativus*), melon (*C. melo*) and watermelon (*Citrullus lanatus*), as well as squashes and pumpkins (*Cucurbita spp.*); all of these being characterized by the development of unisexual male and female flowers on the same plant foot. It is the distribution pattern of male and female flowers on a plant which indicates the different cucurbit sex morphotypes, which are: both female and male flowers (monoecy), hermaphrodite and male flowers (andromonoecy), female, hermaphrodite and male flowers (trimonoecy), only male flowers (androecy), only female flowers (gynoecy) (Galun, 1962;

Malepszy and Niemirowicz-Szczytt, 1991; Perl-Treves, 2004). These sex morphotypes are being used to elucidate the genetic and molecular mechanisms underlying sex determination in this group of species.

It is well established that the plant hormone ethylene is the main regulator of sex determination in cucurbits (Manzano et al., 2010a, 2011; Pannell, 2017; Perl-Treves, 2004). This hormone determines the sex of each floral meristem by arresting the development of one of the sex floral organ primordia. A high production of ethylene is required to arrest the development of stamen primordia, which latter transform the meristem into a female flower, while low production of ethylene arrests the development of carpel primordia, which transform the meristem into a male flower (Kater et al., 2001; Malepszy and Niemirowicz-Szczytt, 1991). Several genes have been identified that control sex morphotypes in cucurbit species, but the complete molecular regulation is still unclear. Most of the genes are specific members of ACC synthase (*ACS*) or ACC oxidase (*ACO*) gene families, both involved in the biosynthesis of ethylene (Yang and Hoffman, 1984). The orthologs *CmACS7*, *CsACS2*, *CitACS4* and *CpACS27* control the arrest of stamen development in cucumber, melon, watermelon and squash respectively, (Boualem et al., 2008, 2009; Ji et al., 2016; Manzano et al., 2016; Martínez et al., 2014) and their mutations result in the conversion of monoecy into andromonoecy. On the other hand, loss of function (LOF) mutations in the orthologs *CsACS11* and *CmACS11* of cucumber and melon, and the *CsACO2* of cucumber completely block the female flower development pathway, leading to androecy (Boualem et al., 2015; Chen et al., 2016). In addition to *ACS* and *ACO* genes, the transcription factor *CmWIP1* has been reported as an essential regulator of male flower development, and its LOF mutation leads to gynoecy. Not only can *WIP1* repress the expression of the *ACS2/7* female gene, but it is also downregulated by *ACS11* in cucumber and melon (Boualem et al., 2015; Hu et al., 2017; Martin et al., 2009).

The role of ethylene signaling components in sex determination is virtually unknown, as none of the sex morphotypes detected so far appear to be caused by changes in the ethylene perception and signaling genes. The evidence that highlights the importance of the ETR1 ethylene receptor in the development of the female flower comes from transgenic experiments in heterologous systems, for example where the cucumber gene *CsETR1* has been overexpressed in Arabidopsis (Wang et al., 2010a), or the *etr1-1* dominant mutant gene of Arabidopsis has been expressed either constitutively or in a specific floral organ of melon (Little et al., 2007). For a better understanding of the ethylene-signaling network regulating

sex determination in cucurbits, we exploited a large mutant platform of *C. pepo* to identify ethylene-insensitive mutants (García et al., 2018). Two of these mutations affect the ethylene receptor genes *CpETR1A* and *CpETR2B* and confer ethylene-insensitivity, plus a subsequent conversion of monoecy into andromonoecy (García et al., 2019).

In this paper we discovered two novel ethylene receptor mutations that are able to confer complete androecy concomitantly with a strong reduction in ethylene sensitivity, and investigated how these ethylene receptor genes interact with each other, as well as with ethylene biosynthesis genes (*ACS11*, *ACS27*, *ACO2*) and with *WIP1*, in the control of sex determination in *C. pepo*. A modified gene-regulating network is proposed, that integrates both ethylene biosynthesis and ethylene perception into the control mechanism of sex determination in cucurbit species.

5.3. Materials and Methods

5.3.1. Plant Material

The ethylene insensitive mutant families *etr1a-1* and *etr1b* were derived from a high throughput screening of a *C. pepo* mutant collection in the search for ethylene triple response of etiolated seedlings (García et al., 2018). Before final phenotyping, each ethylene-insensitive mutant was crossed, for at least three generations, with the background genotype MUC16. Since the two *etr* mutations confer androecy, the homozyous mutants could not be maintained by selfing, but were always selected from segregated populations.

The semi-dominant mutation *etr1b* was maintained in heterozygous and homozygous conditions by selfing the andromonoecious wt/*etr1b* plants in BC_3 and successive BC generations. The dominant mutation *etr1a-1* was always maintained in heterozygosity by backcrossing the androecious wt/*etr1a-1* plants in each BC generation with the background line MUC16.

5.3.2. Phenotyping for ethylene triple response and sex determination traits

The BC_nS_1 and BC_n segregated populations for *etr1b* and for *etr1a-1* mutations were firstly classified by their triple response to ethylene in WT, intermediate (wt/*etr*) and ethylene insensitive mutants *(etr/etr)*, and then transplanted to a greenhouse. There, they were grown to maturity under the standard local conditions in Almería, Spain. The triple response assay was performed using etiolated seedlings as described in García et al., 2019. A minimum of 30 WT and mutant plants were phenotyped for each treatment. The hypocotyl length of

plants grown in air in 50 ppm of ethylene was used to assess the percentage of ethylene sensitivity of each plant and genotype. The sex phenotype of each plant was determined by scoring the sex of the flowers in the first 40 nodes of each plant. We assessed the sex expression of each genotype by determining the node at which plants transited to pistillate flowering (Pistillate Flowering Transition), and the number of nodes with male or pistillate flowers.

The sex phenotype of each individual pistillate flower was assessed by the Andromonoecy Index (Manzano et al., 2016; Martínez et al., 2014). Pistillate flowers were separated into three phenotype classes, that were scored from 1 to 3 according to their degree of stamen development: i) female (AI=1), showing no stamen development; ii) bisexual (AI=2), showing a partial development of stamens and no pollen, ii) hermaphrodite flowers (AI=3), showing a complete development of stamens and pollen. The average AI of each plant and genotype was then assessed from the AI score of at least 5 individual pistillate flowers from each plant, and a minimum of 30 plants for each genotype.

The growth rate of the ovaries and petals in the flowers was assessed by measuring their length every 3 days. At least 12 flowers of each genotype were used, starting with flowers of about 2 mm in length. The anthesis time was measured as the number of days that a 2 mm floral bud takes to reach anthesis. The effect of *etr1b* and *etr1a-1* mutations on plant vigor was assessed by the average plant height, the number of nodes, and the internode lengths in the main shoot of WT and mutant plants, 60 days after sowing.

5.3.3. Identification of *etr1b* and *etr1a-1* mutations by WGS analysis

To identify the causal mutations of *etr1b* and *etr1a-1* phenotypes, WT and mutant plants derived from segregating populations were subjected to whole-genome resequencing (WGS). 100 seedlings from either BC_2S_1 or BC_2 populations, segregating for *etr1b* and *etr1a-1* respectively, were subjected to ethylene triple response treatment. The plants exhibiting the ethylene-insensitive mutant phenotype were separated from the WT ones. The sex phenotypes of both ethylene-sensitive and -insensitive seedlings were assessed in the adult plants so as to confirm the genotype of the plants.

In the *etr1b* segregated population, the ethylene-sensitive and monoecious plants were genotyped as wt/wt, the intermediate ethylene-insensitive and andromonoecious plants were genotyped as wt/*etr1b*, and the ethylene-insensitive and androecious plants were considered as *etr1b/etr1b*. In the *etr1a-1* segregated population, ethylene sensitive and monoecius plants

were genotyped as wt/wt, and ethylene insensitive and androecious plants as wt/*etr1a-1*. The DNA from each individual plant was isolated by using the GeneJET Genomic DNA Purification Kit (Thermo Fisher). DNA from 20 plants having the same phenotype in each mutant family were bulked together. Two DNA bulks were generated for each mutant family: a WT monoecious bulk (wt/wt), and a mutant androecious bulk (*etr1b/etr1b* or wt/*etr1a-1*).

The genomic DNA of each sample was ramdomly sheared into short fragments of about 350 bp for library construction using the NEBNext® DNA Library Prep Kit. Following end-repairing, dA-tailing, and further ligation with the NEBNext adapter, the fragments were briefly PCR enriched with indexed oligos. Pair-end sequencing was performed using the Illumina sequencing platform, with a read length of PE150 bp at each end. The effective sequencing data was aligned with the reference *Cucurbita pepo* genome v4.1 through BWA (Li and Durbin, 2009) software. The duplicates were removed by SAMTOOLS (Li et al., 2009a). Single nucleotide polymorphisms (SNPs) were detected by using the GATK HaplotypeCaller (DePristo et al., 2011). ANNOVAR was used to do annotate the detected SNPs (Wang et al., 2010b).

Common variants between these two ethylene-insensitive mutant families (and other sequenced mutant families in the laboratory) were discarded as they may represent spurious SNPs due to sequencing and/or alignment errors or common genomic differences with the reference genome. Based on the segregation analysis of each mutant phenotype, the genotypes of WT plants should be 0/0, as in the reference genome. As discussed above, the mutant bulks need to be *etr1b*/*etr1b* (1/1) or wt/*etr1a-1* (0/1). Therefore, we expected a mutant allele frequency (AF) of 0 in each WT bulk, an AF=1 in the *etr1b* bulk, and AF=0,5 in the *etr1a-1* bulk.

Data was then filtered according to the following parameters: i) GQ (Genotype Quality) ≥20, ii) DP (read depth) ≥10 (specific base covered with at least 10 reads). The Reference allele frequency was calculated for each SNP position as: Ref. Freq = (Ref. allele depth of andro-bulk + Ref allele depth of WT-bulk) / Total read depth for both bulks. We then filtered based on reference allele frequency (Ref. Freq). This option removes SNPs that for some reason or other are over- or under-represented in both bulks. SNPs with a Reference allele frequency of more than 0.3 and less than 0.7 were filtered out. To identify the candidate mutations of ethylene insensitive and androecious phenotypes, the SNP index was calculated for all the SNP positions in each sequenced bulk as: SNP index = Alternate allele depth / Total read

depth. After that, the derived Δ (SNP index), which combines SNP index information from the two DNA bulks: Δ (SNP index) = (SNP index of WT bulk) - (SNP index of mutant bulk), was also calculated. For a specific location, Δ (SNP index) =1 if the bulked mutant DNA comprises only mutant candidate variants, Δ (SNP index) =−1 if it contains only WT variants, and Δ (SNP index) =0 if both bulks have WT and mutant variants and the same SNP index (Mansfeld and Grumet, 2018; Takagi et al., 2013). Candidate causal mutations for *etr1b* were selected from those variable positions with a Δ (SNP index) =0.9-1, and candidate causal mutations for *etr1a-1* from positions with a Δ (SNP index) =0.4-0.6.

5.3.4. Validation of the identified mutations by high-throughput genotyping of individual segregating plants

A segregation analysis was performed to confirm that the identified mutations were the causal mutations of the *etr1b* and *etr1a-1* phenotypes. About 200 plants from BC_2S_1 and BC_3S_1 populations segregating for *etr1b*, together with BC_3 and BC_4 populations segregating for *etr1a-1*, were genotyped by using real time PCR with TaqMan probes. The multiplex PCR reactions were done with the Bioline SensiFAST™ Probe No-ROX Kit -a set of forward and reverse primers amplifying the polymorphic sequence- and two allele-specific probes descriptive of the SNP of interest (C-T). The WT probe was labelled with FAM dye, while the mutant probe was labelled with HEX reporter dye. The BHQ1 quencher molecule was used in both probes (Supplementary Table 5.1). The annealing temperature in the reaction was 64ºC for *etr1b*, and 60ºC for *etr1a-1*. After genotyping, adult plants were phenotyped for ethylene triple response and for sex expression, so as to discover whether the mutant alleles co-segregated with the *etr1b* mutant phenotype or with the *etr1a-1* one

5.3.5. Relative gene expression assessment by Quantitative RT-PCR

Gene expression analysis was carried out using WT and mutant plants which were grown in a culture chamber at 25 ºC during a 16/8 h day/night photoperiod. The expression level was studied in the apical shoots of plants at two developmental stages: M0 and M1. M0 samples correspond to apical meristems from plants having 2-4 extended leaves, and M1 to apical meristems of plants having 10-12 leaves. At M0, female flowering transition had still not occurred. Although both WT (monoecious) and mutant (androecious) apical shoots only bear male floral buds, at M1 the WT monoecious plants had already transited to female flowering, while the mutant androecious plants were still producing male flowers. Therefore, at M1, WT apical shoots hold both male and female floral buds, while mutant plants have male floral buds only.

Expression levels were assessed according to quantitative RT-PCR in the genes encoding for ethylene receptors (*CpETR1A*, *CpETR1B* and *CpETR1A*, *CpETR2B*), ethylene biosynthesis (*CpACS27*, *CpACS11* and *CpACO2*), and *WIP1*. Three biological replicates were analyzed for each sample. RNA was isolated according to the protocol of the GeneJET Plant RNA Purification Kit (Thermo Fisher). cDNA was then synthesized by using RevertAid RT Reverse Transcription Kit (Thermo Fisher). The expression levels of genes were also evaluated through qRT-PCR, by using the Rotor-Gene Q thermocycler (Qiagen) and the SYBR® Green Master Mix (BioRad). Supplementary Table 5.2 shows the primers used in qRT-PCR reactions for each analyzed gene.

5.3.6. Bioinformatics and statistical analyses

Alignments were performed using the BLAST alignment tools at NCBI (http://www.blast.ncbi.nlm.nih.gov/) and Clustal Omega at EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/). The phylogenetic relations of ACS, ACO and WIP1 genes were studied using MEGA7 software (Kumar et al., 2016). This allowed for the alignment of proteins and the construction of phylogenetic trees using the Neighbor-Joining method (Saitou and Nei, 1987) to take place. The evolutionary distances were computed using the JTT matrix-based method (Jones et al., 1992), with 1000 bootstrap replicates. The protein sequences (Supplementary Table 5.3) were obtained using The Arabidopsis Information Resource (https://www.arabidopsis.org/) and the Cucurbit Genomics Database (http://cucurbitgenomics.org/).

Multiple data comparisons were obtained by the analysis of variance with the significance level $P<0.05$, and each two averages were compared using Fisher's least significant difference method.

5.4. Results

5.4.1. Both *etr1b* and *etr1a-1* ethylene-insensitive mutations confer androecy

The *etr1b* and *etr1a-1* mutants were isolated from an EMS-mutagenized M_2 collection - after screening for ethylene triple-response in etiolate seedlings - and were found to correspond to the so-called *ein1* and *EIN4* mutants in (García et al., 2018). As the two mutants are female sterile, they were maintained in segregated populations while their inheritance patterns were studied (Supplementary Table 5.4). The *etr1b* is a semi-dominant mutation that was maintained by selfing heterozygous wt/*etr1b* plants from backcrossing populations (BC_nS_1 generations). The *etr1a-1* is a dominant mutation only maintained in heterozygosity after

backcrossing with the background genotype MUC16 for a number of generations (BC_n generations). Figure 1 shows the triple response phenotypes of each of the mutants. Homozygous wt/wt seedlings respond to ethylene by reducing the length of their hypocotyl and roots, and by increasing the thickness of their hypocotyl. In contrast, homozygous *etr1b/etr1b* and heterozygous wt/*etr1a-1* seedlings barely responded to ethylene, showing similar hypocotyl and root lengths to seedlings grown in air. Heterozygous wt/*etr1b* seedlings showed an intermediate triple-response phenotype compared to the two homozygous ones (Figure 1).

Assuming that WT plants are completely sensitive to ethylene (0% ethylene insensitivity), we estimated the percentage of ethylene insensitivity in homozygous (*etr1b/etr1b*) and heterozygous (wt/*etr1b* and wt/*etr1a-1*) seedlings for the two mutations. The heterozygous plants of *etr1a-1* showed the most severe ethylene-insensitive phenotype (99.6% ethylene insensitivity). The homozygous and heterozygous plants of *etr1b* showed 88.5 and 50.6% of ethylene insensitivity, respectively (Figure 1).

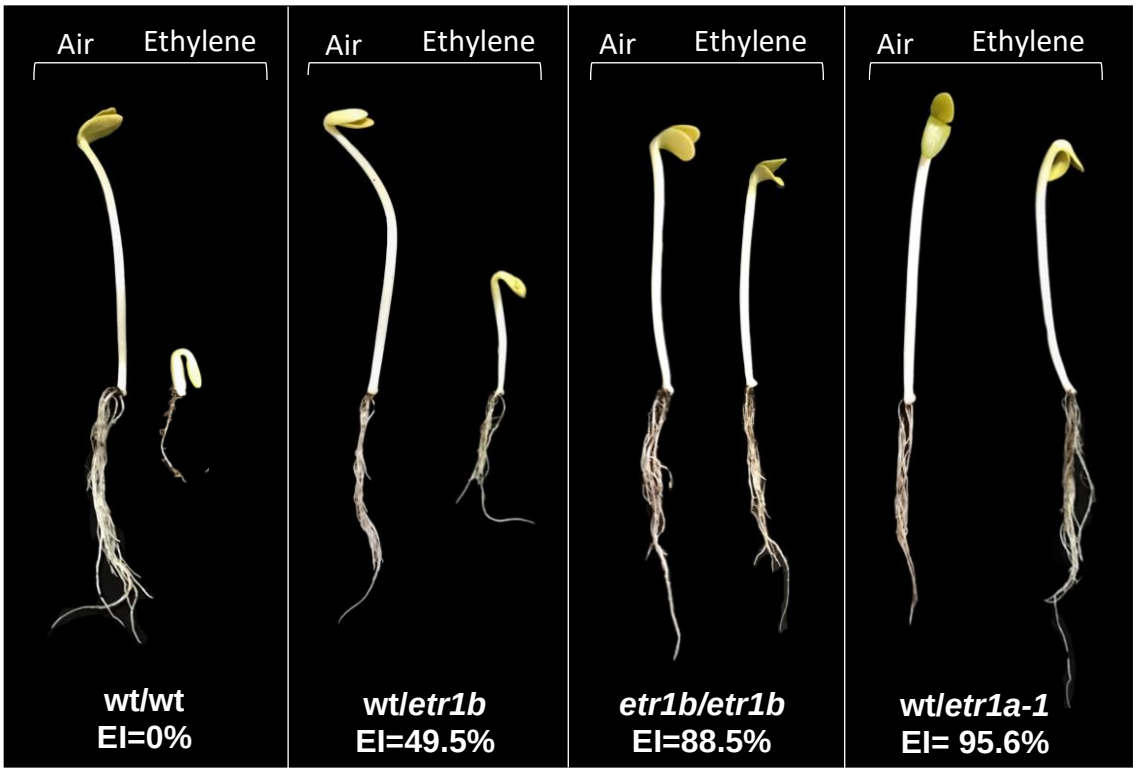


Figure 1. Ethylene triple-response phenotypes of WT, heterozygous and homozygous plants for *etr1b* and *etr1a-1* mutations. The average level of ethylene insensitivity (EI) for each mutant was assessed as the percentage of reduction in hypocotyl length with respect to air-grown seedlings, assuming that ethylene insensitivity in WT plants is 0%. n=20 for each genotype.

Given that ethylene is the main regulator of cucurbit sex determination, we assessed the sexual phenotype of *etr1b* and *etr1a-1* mutants. Staminate and pistillate flowers were recorded in the first 40 nodes of WT, heterozygous and homozygous mutant plants; flowers and plants were classified according to the andromonoecious index (AI) (García et al., 2019). WT plants were all monoecious and showed two sexual phases of development: a first phase in which plants produced only male flowers, and a second phase where plants alternated the production of male and female flowers. WT female flowers showed a complete arrest of stamen development (AI=1) (Figure 2). Heterozygous wt/*etr1b* were partially andromonoecious, showing a partial conversion of female flowers into bisexual or hermaphrodite flowers (average AI=1.8; Figure 2). The female flowering transition period was delayed in wt/*etr1b* plants; the number of male flowers per plant increased. The mutants with the most extreme ethylene-insensitive phenotype (*etr1b/etr1b* and wt*/etr1a-1*) were all androecious and only bore male flowers on the main shoot (Figure 2). This data demonstrates that the initiation of female flower development requires the plant to be sensitive to ethylene. The partial ethylene insensitivity of heterozygous wt/*etr1b* mutants disrupts the stamen arrest of female flower production, inducing a conversion of female into bisexual or hermaphrodite flowers, and also converts monoecy into partial andromonoecy. On the other hand, the almost complete ethylene insensitivity of homozygous *etr1b*/*etr1b* and heterozygous wt/*etr1a-1* mutants prevented the development of carpels, which resulted in a complete conversion of female into male flowers and of monoecy into androecy.

5.4.2. Both plant vigor as well as petal and fruit development are modified by *etr1b* and *etr1a-1*

The vegetative development of adult *etr1b/etr1b* and wt/*etr1a-1* plants was found to be enhanced. Table 1 shows the differences in plant height and node length of WT and mutant plants grown under the same conditions. The two mutations increased the growth rate of plants, the two mutants showing greater height than WT plants. These differences in height were due to an increase in the internode length, whose average was twice as high in mutants as in WT (Table 1).

The *etr1b* and *etr1a-1* mutations also modified the development and level of maturity of both male and female flowers. The growth rate of WT petals in female flowers was much higher than that of male flowers. This causes the former to mature and reach anthesis approximately one week earlier than the latter (Figure 3).

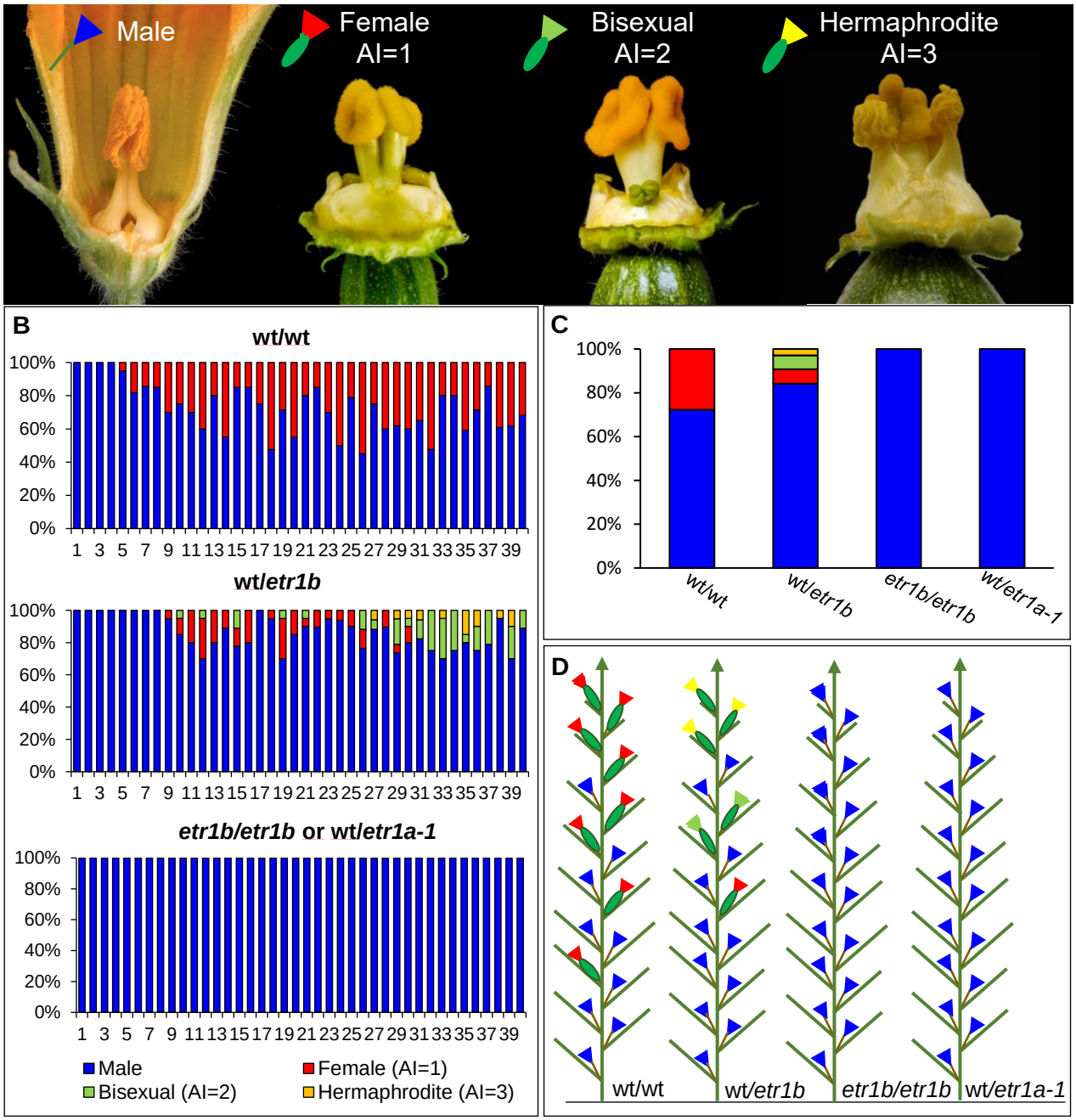


Figure 2. The effect of *etr1a-1* and *etr1b* mutations on sex determination in *C. pepo*. (**A**) Phenotypes of male, female, bisexual and hermaphrodite flowers. The andromonoecious index (AI) is indicated for each pistillate flower (AI=1, complete arrest of stamen development; AI=2, partial arrest; AI=3, no arrest). (**B**) Distribution of the different types of flowers in the 40 first nodes of WT plus *etr1a-1* and *etr1b* mutant plants. In each node the colored bars indicate the percentages of the total number of plants analyzed: blue = male, red = female, green = bisexual, yellow = hermaphrodite. N=30 plants for each genotype. (**C**) Comparison of the percentage of male, female, bisexual and hermaphrodite flowers in WT mutants and those of *etr1a-1* and *etr1b*. (**D**) Schematic representation of the distribution of male and female flowers in WT and *etr1a-1* or *etr1b* mutant plants. Blue, male flower; red, female; green, bisexual; yellow, hermaphrodite.

Table 1. Effects of *etr1b* and *etr1a-1* mutations on plant vegetative development.

Mutant	Genotype	Average plant height (cm)	Average node length (mm)
etr1b	wt/wt	74.7c	12.5c
	etr1b/etr1b	124.8b	23.2b
etr1a-1	wt/wt	84.9c	14.6c
	wt/*etr1a-1*	175.3a	30.2a

Different letters indicate statistically significant differences between WT and mutant genotypes of the same family (P<0.05).

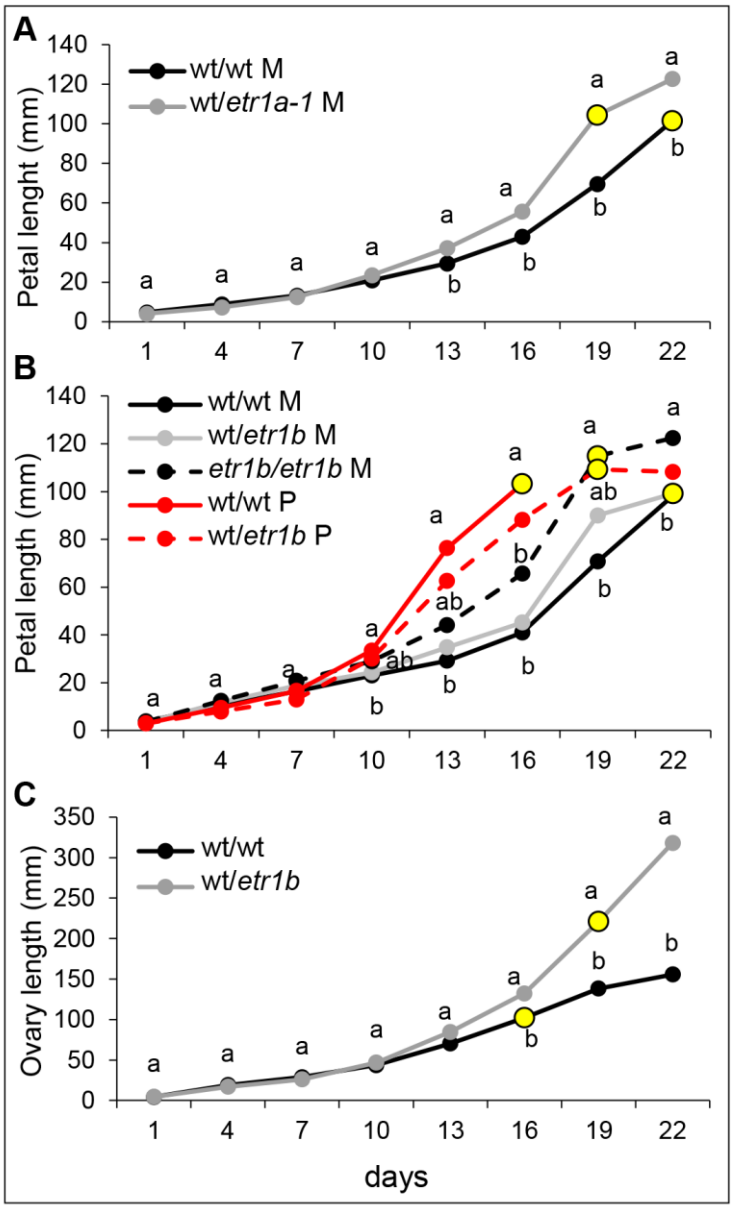


Figure 3. (**A, B**) Effect of *etr1a-1* and *etr1b* mutations on the growth rate and maturation of petals in male (M; black and grey lines) and pistillate flowers (P; red lines). Yellow circles indicate the time at which more than 80% of the flowers reach anthesis. (**C**) Effect of *etr1b* on the growth rate of ovaries in pistillate flowers.

In male flowers, *etr1b* and *etr1a-1* enhanced the growth rate of petals, allowing them to reach anthesis earlier (Figure 3). In bisexual and hermaphrodite flowers of the heterozygous mutant wt/*etr1b,* petal growth rate was reduced compared to that of WT female flowers, thus delaying both maturation and anthesis in mutant pistillate flowers (Figure 3). Anthesis time -measured as the time that a 2 mm floral bud takes to reach anthesis and to open -was longer in male flowers (average 22 days) than in female WT ones (average 16 days). Hermaphrodite *etr1b* flowers, however, took an average of 19 days to anthesis. Also, under our greenhouse conditions, several hermaphrodite flowers never even reached anthesis. Their petals remained green and closed for more than 30 days.

Significant differences in ovary size were also detected among wt/wt and wt/*etr1b* pistillate flowers (Figure 3C). While unpollinated WT ovaries reached about 10 cm in anthesis and then aborted, those of wt/*etr1b* flowers reached 20-40 cm at anthesis (Figure 3C). These can be considered parthenocarpic fruits, since the flowers were closed, and were therefore not available for pollination. During the first 13 days, the growth rates of WT and mutant ovaries were similar. After 16 days, unfertilized WT ovaries started to abort, while those of heterozygous wt/*etr1b* flowers maintained their growth rate (Figure 3C). Since anthesis time of wt/*etr1b* pistillate flowers was achieved much later than in WT, the size of the mutant ovary at anthesis was much bigger (Figure 3C).

5.4.3. *The etr1a-1* and *etr1b* mutations affect the transmembrane domains of the ethylene receptors CpETR1A and CpETR1B

Searching for the causal mutation of *etr1a-1* and *etr1b* phenotypes, we performed a whole-genome re-sequencing of two bulked DNA samples, one containing DNA from 20 WT/WT plants (ethylene-sensitive and monoecious), and the other containing 20 mutant plants (ethylene-insensitive and androecious) in the segregating populations for each mutation, which corresponded to either *etr1b*/*etr1b* and wt/*etr1a-1* genotypes. The bioinformatic analysis of sequencing data resulted in the identification of single-nucleotide polymorphisms (SNPs) in monoecious and androecious phenotypes in both EMS mutant families. The identified SNPs were filtered by using different criteria (Supplementary Table 5.5). Common variants between samples of the two mutants were discarded as they probably correspond to spontaneous nucleotide polymorphisms in the MUC16 genetic background (García et al., 2018). SNPs corresponding to canonical EMS mutations (C>T and G>A) (Drake and Baltz, 1976) were then selected, and filtered according to their quality and depth, as well as to the SNP-index in WT and mutant samples (Supplementary Table 5.5). Variants having a SNP-

index of 0 in the WT sample and 0.9-1 in the *etr1b* sample (Δ SNP index =0.9-1) were selected as candidate mutations for *etr1b* phenotype, while those having a SNP-index of 0 in WT sample and 0.4-0.6 in the *etr1a-1* sample (Δ SNP index =0.4-0.6) were selected as the candidate mutations for *etr1a-1* (Figure 4A). One SNP in chromosome 11, and 2 SNPs in chromosome 7, met the established criteria for *etr1b* and *etr1a-1* mutations, respectively (Figure 4A). Among them, only two C>T transitions in the coding region of ethylene receptor genes *CpETR1B* (in chromosome 11) and *CpETR1a* (in chromosome 7) were finally selected as the putative causal mutations of *etr1b* and *etr1a-1* phenotypes respectively, because they have a high impact on the proteins (Supplementary Table 5.5). The other *etr1a-1* putative causal mutation was discarded because it does not co-segregate with the phenotype.

To confirm that the selected EMS mutations were the ones responsible for *etr1b* and *etr1a-1* phenotypes, we performed a segregating analysis, searching for mutations and phenotypes in hundreds of plants from BC_3S_1 and BC_3 segregating populations (Table 2). The ethylene insensitivity and the androecious phenotypes of *etr1b* and etr*1a-1* co-segregated perfectly with the mutations C>T in the coding region of the genes *CpETR1B* and *CpETR1A* respectively. From a population of 287 plants segregating for *etr1b*, all monoecious plants (74) were homozygous for the reference allele (0/0) while all the andromonoecious (87) and androecious (126) plants were heterozygous (0/1) or homozygous (1/1) segregating for the alternative mutant allele. In the segregating populations for *etr1a-1* (157 total plants), all the monoecious plants were also homozygous for the WT allele (0/0), while the androecious ones were all heterozygous for the mutation (1/0). This data demonstrates that the C>T EMS mutations detected in *CpETR1B* and *CpETR1A* genes co-segregate 100% with the mutant phenotypes, and therefore that these are likely to be the causal mutations for *etr1b* and *etr1a-1* phenotypes respectively.

Given that the genomes of *Cucurbita* species are duplicated (Montero-Pau et al., 2017; Sun et al., 2017), *CpETR1A* and *CpETR1B* correspond to two *ETR1* paralogs of the *C. pepo* genome which showed more than 90% of homology. The *etr1a-1* and *etr1b* mutations are both located in the first exon of each of the genes (positions 281 and 107 respectively) (Figure 4C). The deduced CpETR1A and ETR1B receptors comprise the domains of other plant ethylene receptors: A N-terminal transmembrane domain sensor, composed of three transmembrane segments, a GAF domain, one histidine-kinase domain, and a response regulator receiver domain (Figure 4C). The mutant *etr1a-1* is an amino acid substitution of

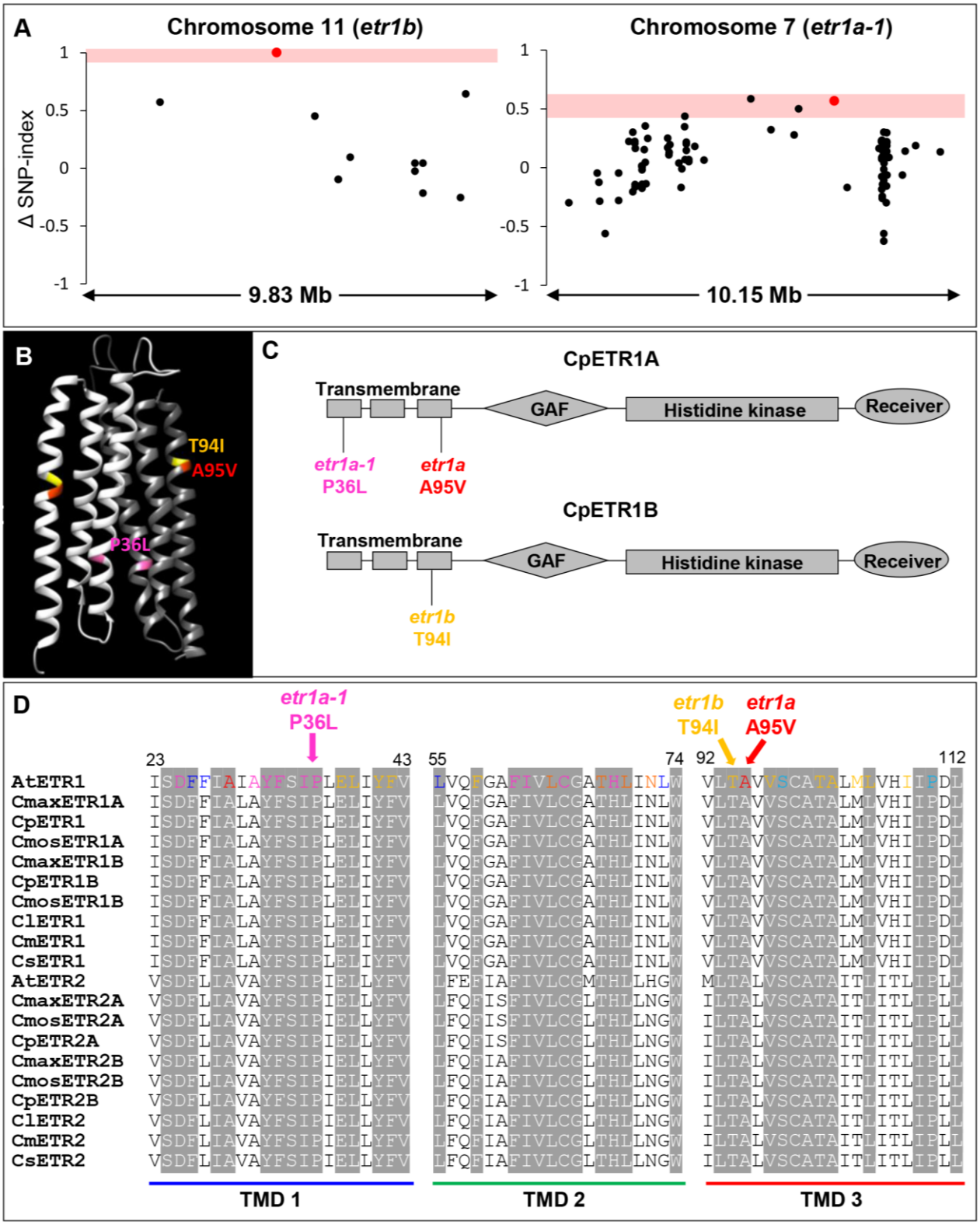
proline by leucine at residue 36 (P36L) in the first transmembrane domain of CpETR1A, while *etr1b* produces an amino acid change of threonine to isoleucine at residue 94 (T94I) in the third transmembrane domain of CpETR1B (Figure 4B, D). Both amino acid substitutions are conserved in the ETR1 receptors of very distinct plant species, which indicates that the residues are functionally important regarding ethylene reception and/or signal transmission (Figure 4D; Supplementary Figure 5.1).

Table 2. Co-segregation analysis of the identified mutations in *CpETR1B* and *CpETR1A* with the ethylene insensitive phenotypes of *etr1b* and *etr1a-1*.

Segregating population	Genotypes	Triple response to ethylene/Sex phenotype			Total No plants
		Ethylene Sensitive/ Monoecious	Intermediate/ Andromonoecious	Ethylene insensitive/ Androecious	
etr1b	0/0	74	-	-	**287**
	0/1	-	87	-	
	1/1	-	-	126	
etr1a-1	0/0	86	-	-	**157**
	0/1	-	-	71	

Genotypes: 0, reference genome WT allele; 1, mutant allele.

Figure 4. Molecular identification and characterization of *etr1b* and *etr1a-1*. (**A**) Δ (SNP-index) graph of each SNP position on the chromosomes 11 and 7, calculated by subtracting the mutant allele frequency (SNP index) of the WT bulk from that of the mutant bulk. SNPs that met the conditions to be considered as causal mutations of the *etr1b* phenotype were located on chromosome 11(pink shadow), while those of the *etr1a-1* phenotype were located on chromosome 7 (pink shadow). The confirmed causal mutations of the two ethylene-insensitive phenotypes are indicated in red. (**B**) 3D model of the transmembrane domain of ETR1 dimers showing the position of each amino acid change. (**C**) Domain structure of the CpETR1A and CpETR1B ethylene receptors, showing the positions of *etr1b* and *etr1a-1* mutations characterized in this paper, as well as *etr1a*, a previously identified mutation in *CpETR1A* (García et al., 2019). (**D**) Multiple alignment of the three transmembrane regions (TMD1, TMD2, and TMD3) of ETR1 and ETR2 indicating the conserved amino acids (grey) in Cucurbits and Arabidopsis. The different colors used indicate the function of each amino acid according to Schott-Verdugo et al. (2019). Pink, ethylene binding; yellow, signal transmission; blue, on-state residues. (Next page)


A
Chromosome 11 (etr1b)
Chromosome 7 (etr1a-1)
Δ SNP-index
1
0.5
0
-0.5
-1
9.83 Mb
10.15 Mb
B
T94I
A95V
P36L
C
CpETR1A
Transmembrane
GAF
Histidine kinase
Receiver
etr1a-1
P36L
etr1a
A95V
CpETR1B
Transmembrane
GAF
Histidine kinase
Receiver
etr1b
T94I
D
etr1a-1
P36L
etr1b
T94I
etr1a
A95V
23
43
55
74
92
112
AtETR1
CmaxETR1A
CpETR1
CmosETR1A
CmaxETR1B
CpETR1B
CmosETR1B
ClETR1
CmETR1
CsETR1
AtETR2
CmaxETR2A
CmosETR2A
CpETR2A
CmaxETR2B
CmosETR2B
CpETR2B
ClETR2
CmETR2
CsETR2
TMD 1
TMD 2
TMD 3

5.4.4. Interactions between ethylene receptor genes

The interaction between ethylene receptor mutations was studied in double mutants for three ethylene receptor genes: *CpETR1A*, *CpETR1B* and *CpETR2B*. We combined *etr1a-1* with *etr1b*, and also *etr1a* with *etr2b*, two previously identified mutations that resulted in a less severe ethylene-insensitive phenotype and into andromonoecy in *C. pepo* (Table 3, (García et al., 2019)). Given that all the *etr* mutations lead to andromonoecy or androecy, and therefore to female sterility, double mutants are unlikely to be generated. Double mutants in all possible combinations were therefore obtained and characterized under heterozygous conditions (Table 3).

Table 3. Ethylene insensitivity and sexual phenotypes of single and double ethylene insensitive mutants.

	Ethylene insensitivity (%)	No. Male flowers[a]	No. Pistillate flowers[a]	Andromonoecy Index	Pistillate flowering transition[b]	Sexual phenotype	References
wt/wt	0e	28.9e	11.2a	1d	7.1d	Monoecious	
wt/*etr1a*	51.1d	32.7e	7.3b	1.1d	10.9cd	Monoecious	García et al., 2019
etr1a/etr1a	85.5b	37.3b	2.6de	2.9a	24.8a	Andromonoecious	García et al., 2019
wt/*etr2b*	43.5d	30.2e	9.9a	1.1d	8.6cd	Monoecious	García et al., 2019
etr2b/etr2b	63.7c	33.7cd	6.3bc	1.8c	11.8c	Partially andromonoecious	García et al., 2019
wt/*etr1b*	50.6d	32.7e	7.3b	1.8c	10.9cd	Partially andromonoecious	This paper
etr1b/etr1b	88.5ab	40a	-	-	no transition	Androecious	This paper
wt/*etr1a-1*	95.6a	40a	-	-	no transition	Androecious	This paper
wt/*etr2b*; wt/*etr1a*	61.3c	35.6bc	4.4cd	2.3b	17.9b	Andromonoecious	García et al., 2019
wt/*etr2b*; wt/*etr1b*	81.2b	37.8ab	2de	3a	9.8cd	Andromonoecious	This paper
wt/*etr1a*; wt/*etr1b*	88.2ab	40a	-	-	no transition	Androecious	This paper
wt/etr2b; wt/*etr1a-1*	90.1a	40a	-	-	no transition	Androecious	This paper
wt/*etr1b*; wt/*etr1a-1*	92.2a	40a	-	-	no transition	Androecious	This paper

The numbers are the average of a minimum of 20 plants for each genotype. [a]Flowers were recorded in the first 40 nodes of each plant. [b]Number of male nodes before female flowering transition. Different letters within the same column indicate significant differences at $P<0,05$.

Despite this, double heterozygous mutants showed very severe ethylene-insensitive phenotypes in the triple-response test: 81.2% ethylene insensitivity in wt/*etr2b* wt/etr1b,

88.2% in wt/*etr1a* wt/*etr1*, 90.1% in wt/*etr2b* wt/*etr1a-1* and 92.2% in wt/*etr1b* wt/*etr1a-1* (Table 3). Each individual mutation had an additive effect on the ethylene insensitivity phenotype of double-mutants, showing the double-heterozygous mutant to be a phenotype that is similar to that of the stronger homozygous single mutant (Table 3).

5.4.5. Effects of *etr1b* and *etr1a-1* on the expression of other sex determination genes

Since the mechanisms of sex determination take place in the floral meristem, where it is determined whether the meristem will produce male or female flowers, we studied the sex expression of ethylene biosynthesis and signaling genes in the apical shoots of WT and *etr1b* and *etr1a-1* mutants. These apical shoots contain floral buds at different stages of development. Apical shoots were collected at two stages of development: M0 (plants with 2 leaves), which is when WT and mutant plants are only producing male flowers; M1 (plants with 10-12 leaves), which is when, although monoecious WT plants have already transited to female flowering, the androecious mutants *etr1b* and *etr1a-1* are still producing male flowers.

Gene expression was always higher in M1 than in M0 for all the studied genes (Figure 5), which suggests that the characterized ethylene genes are upregulated during the transition to female flowering. Gene expression of four ethylene receptor genes is shown in Figure 5. Since the genome of *C. pepo* is duplicated, it contains two paralogs for *ETR1* (*CpETR1A* and *CpETR1B*), and two paralogs for *ETR2* (*CpETR2A* and *CpETR2B*). The mutations *etr1a-1* and *etr1b* did not significantly alter the transcription of the ethylene receptor genes, which discard any autoregulation of ethylene receptor-gene expression by ethylene in the apical shoot. The only ethylene receptor gene that was found to be downregulated by *etr1a-1* but not by *etr2b*, was *CpETR2A*, whose expression was reduced in mutant *etr1a-1* plants at M0 and M1 stages of development (Figure 5). The expression of ethylene biosynthesis genes (*ACS27*, *ACS11* and *ACO2*) and the transcription factor gene *WIP1* -known to control sex determination in other cucurbit species- was analyzed.

The *WIP1* genes (*CpWIP1A* and *CpWIP1B*), and the *ACO* and *ACS* genes (*CpACO2A* and *CpACO2B*, *CpACS27A* and *CpACS27B*, and *CpACS11A*), were selected from a number of *WIP1*, *ACO* and *ACS* genes in the genome of *C. pepo* according to their homology and phylogenic relationship with melon and cucumber genes whose function in sex determination is known (Supplementary Figure 5.2; Supplementary Table 5.3).

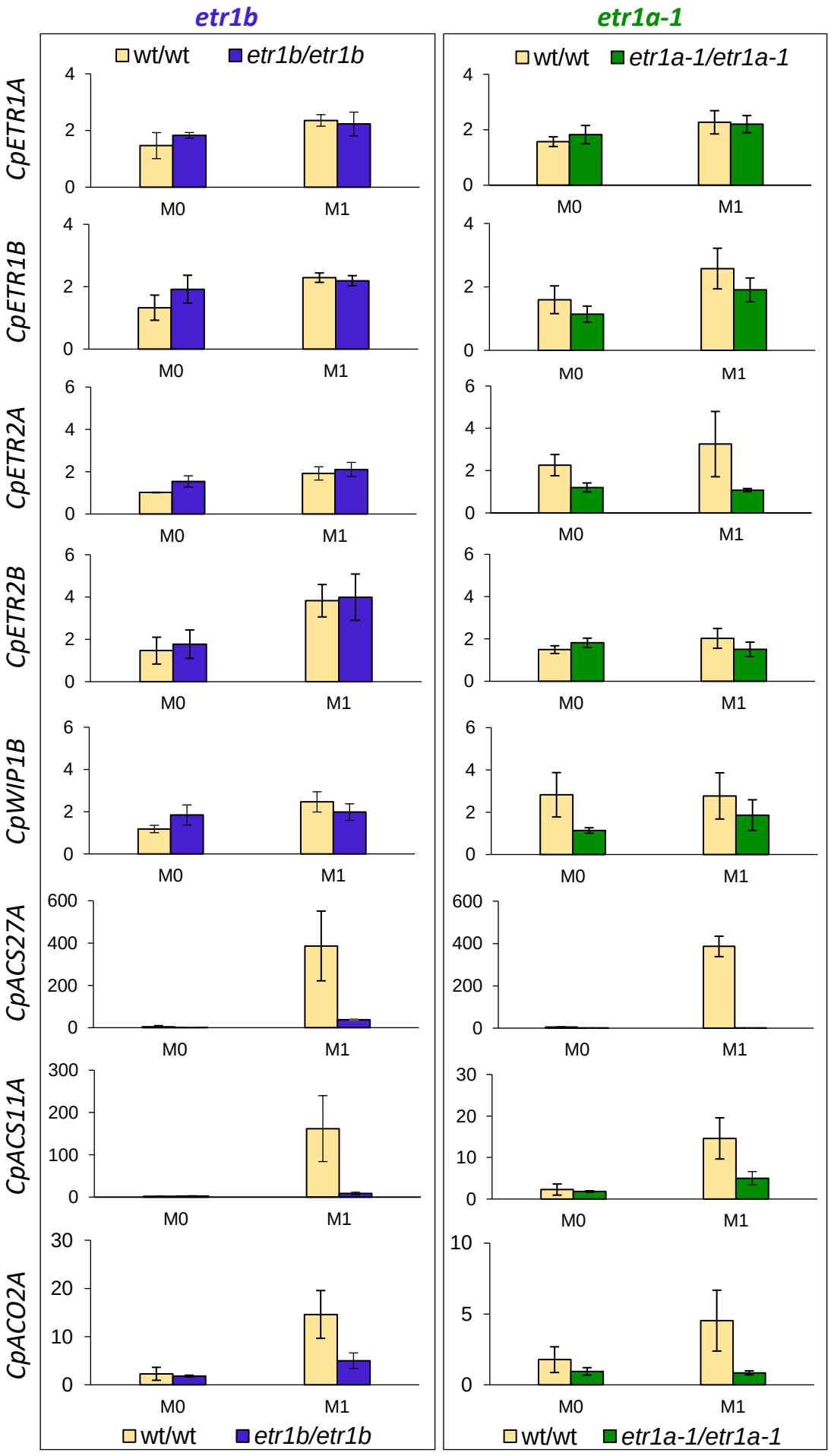

etr1b
etr1a-1
wt/wt
etr1b/etr1b
etr1a-1/etr1a-1
CpETR1A
CpETR1B
CpETR2A
CpETR2B
CpWIP1B
CpACS27A
CpACS11A
CpACO2A
M0
M1

Figure 5. Relative gene expression levels in WT and *etr1b* and *etr1a-1*. The expression levels were studied for ethylene receptor genes *CpETR1* and *CpETR2*, and also for the following genes that are known to be involved in sex determination in cucurbit species: *CpWIP1B*, *CpACS27A*, *CpACS11* and *CpACO2A*. Only the paralogs showing expression levels in the apical shoots are shown. The relative level of each transcript was quantified by qPCR in three independent replicates of each tissue. M0 corresponds to apical shoots of plants having 2-4 extended leaves, and M1 to apical shoots of those having 10-12 leaves. At M1 stage, the female flowering transition had occurred in WT plants (monoecious), but not in the mutant plants (androecious). The gene expression level comparison was made between wt/wt and *etr1b/etr1b* on the one hand, and between wt/wt and wt/*etr1a-1* plants on the other. Error bars represent standard deviation.

Of the two *CpWIP1* paralogs, *CpWIP1A* showed no expression. *CpWIP1B* was faintly downregulated only in *etr1a-1* apical shoots at the M0 stage of development, not at the M1 stage (Figure 5). The expression of *CpWIP1B* was not altered, neither in *etr1b* nor in *etr1a-1* apical shoots at M1 (Figure 5).

Regarding the two paralogs of *ACO2* and *ACS27* detected (Supplementary Figure 5.2; Supplementary Table 5.3), only one of them (*CpACO2A* and *CpACS27A*) was found to be expressed in the apical shoot. For *ACS11*, only one gene, *CpACS11*, was identified in the genome of *C. pepo* (Supplementary Figure 5.2; Supplementary Table 5.3). The expression of ethylene in all these ethylene biosynthesis genes was significantly downregulated in the two ethylene insensitive mutants (Figure 5). At the M0 developmental stage, where the female flowering transition has not yet occurred, the transcription of *CpACS11*, *CpACS27A* and *CpACO2A* was found to be very low both in the WT and mutants. However, in M1, when WT plants had already transited to female flowering, there was an upregulation of all analyzed ethylene biosynthesis genes in the apical shoots of WT, but not in those of *etr1b* and *etr1a-1* mutants, where the expression remained as in the M0 developmental stage (Figure 5). These findings indicate that *CpACS11*, *CpACS27A* and *CpACO2A* are positively autoregulated by the action of ethylene upon female flowering transition.

5.5. Discussion

5.5.1. Missense *etr1a-1* and *etr1b* mutations alter the transmembrane domain of ethylene receptors and lead to ethylene insensitivity

Ethylene is perceived by a family of two-component histidine kinase receptors, which repress ethylene signaling cascade in the absence of ethylene but become inactive upon ethylene binding (Hua and Meyerowitz, 1998). The *C. pepo* genome contains six ethylene receptor genes. Given that a whole-genome duplication occurred in the species of the genus

Cucurbita (Montero-Pau et al., 2017; Sun et al., 2017), these receptor genes correspond to two duplicated paralogs for either *ETR1* (*CpETR1A* and *CpETR1B*), *ETR2* (*CpETR2A* and *CpETR2B*) or *ERS1* (*CpERS1A* and *CpERS1B*), each one being located in the subgenome A or B of *C. pepo.*

In this paper, we report the finding of two newly-discovered ethylene-insensitive mutations (*etr1a-1* and *etr1b*), which alter the transmembrane domain of both CpETR1A and CpETR1B ethylene receptors. This N-terminal domain of ethylene receptors is composed of three membrane-spanning helices that are required for ethylene binding being necessary a copper co-factor (Rodriguez et al., 1999). Some residues in these hydrophobic helices are essential for ethylene binding (Hall et al., 1999; Rodriguez et al., 1999; Schaller et al., 1995). Different *ETR1* mutations in Arabidopsis discern the role of the different segments and amino acids of the transmembrane domains in the regulation of ethylene binding and signal transmission (Wang et al., 2006). Also, a structural model of the transmembrane sensor domain has recently been built (Schott-Verdugo et al., 2019).

The dominant mutation *etr1a-1* is a P36L amino acid exchange that affects the first transmembrane helix of CpETR1A, while the semi-dominant mutation *etr1b* is a T94I amino acid exchange in the third transmembrane helix of CpETR1B. Both P36 and T94 are known to be highly conserved residues in the transmembrane domains of ethylene receptors of many plants (Figure 4, Supplementary Figure 5.1; (Wang et al., 2006)). Their mutations in Arabidopsis ETR1 also confer strongly dominant insensitivity upon ethylene, suppressing the constitutive ethylene response of the *etr1 etr2 ein4* triple mutant. This indicates that due to the actions of these mutations, ETR1 develops into a constitutively active state that cannot be inactivated by ethylene (Wang et al., 2006). The structural model of the transmembrane sensor domain of ETR1 has revealed that T94 is not essential for ethylene binding. It is, however, relevant to signal transmission (Schott-Verdugo et al., 2019). The residue P36 is in close proximity to ethylene and copper binding sites, and is essential for the ethylene binding activity of ETR1 (Schott-Verdugo et al., 2019). Two previously-identified ethylene-insensitive mutations in *C. pepo*, *etr1a* and *etr2b*, are also single amino acid substitutions that disrupt the third transmembrane helix of CpETR1A (A95V) and the coiled-coil domain between the GAF and histidine-kinase domains of CpETR2B (E340K), respectively. Both of these are likely to affect ethylene signal transmission, without affecting ethylene binding activity (García et al., 2019).

The observed differences in the percentage of ethylene insensitivity between the different single homozygous and heterozygous mutants, and between single and double mutants in *CpETR1A*, *CpETR1B* and *CpETR2B* genes (Table 3) are probably associated with the position of the residue within the N-terminal domain of the protein, as occurs in Arabidopsis (Schott-Verdugo et al., 2019). These differences also indicate functional redundancy, and also the cooperation of these three receptors with regard to the perception and signaling of ethylene. The semi-dominant nature of these mutations, and the different levels of ethylene insensitivity in heterozygous double mutants, demonstrate that the final level of ethylene insensitivity of the plant depends not only upon the strength of each mutation, but also on the dosage of WT and mutated alleles for all the ethylene receptor genes (Table 3). The plants that show the highest levels of ethylene insensitivity are those with the two strongest mutated alleles, irrespective of whether they come from the same or from different *ETR* genes (Table 3). Moreover, the fact that *etr1a*, *etr1a-1* and *etr1b* confer ethylene insensitivity, confirms that the function of these two paralogous *ETR1* genes, *CpETR1A* and *CpETR1B*, is conserved after the genome duplication of the *Cucurbita* genus (Sun et al., 2017). The variation in the severity of these mutant phenotypes is probably not due to differences in the functioning of *CpETR1A* and *CpETR1B* paralogs, but to the relevance of the mutated protein residue, as discussed above.

5.5.2. Ethylene receptors control sex determination, and their mutations lead to andromonoecy and androecy

Previous reports indicate that naturally-occurring andromonoecy and androecy in cucurbit species are the result of mutations in certain ethylene biosynthesis genes. Mutations in *ACS2/7* result in andromonoecy, while mutations in *ACS11* and *ACO2* result in androecy (Boualem et al., 2008, Boualem et al., 2009, Li et al., 2009; Martinez et al., 2014; Manzano et al., 2016; Boualem et al., 2015; Chen et al., 2016). However, in this paper we demonstrate that andromonecy or androecy do not result from mutations in different specific ethylene receptor genes, but is conferred by mutations in either *CpETR1A*, *CpETR1B* or *CpETR2B* ethylene receptor genes. The occurrence of one sex type or another depends on the level of ethylene insensitivity conferred by the different mutations (Figure 6). After comparing 12 single- and double-mutant genotypes, we concluded that there is a strong correlation between the percentage of ethylene insensitivity in etiolated seedlings with the final sex phenotype of the plant (Figure 6). Monoecy in the ethylene-sensitive WT plants (0% ethylene insensitivity) is converted into andromonoecy in the partially ethylene-insensitive single and

double mutants, and into androecy in the mutants with the most strongly ethylene-insensitive phenotype. Although the limits between monoecy, andromonoecy and androecy are not very clearly defined, the transition from monoecy to andromonoecy occurs when ethylene insensitivity is increased to more than 50%, while the transition from andromonoecy to androecy occurs when the insensitivity level is increased to more than 88% (Figure 6).

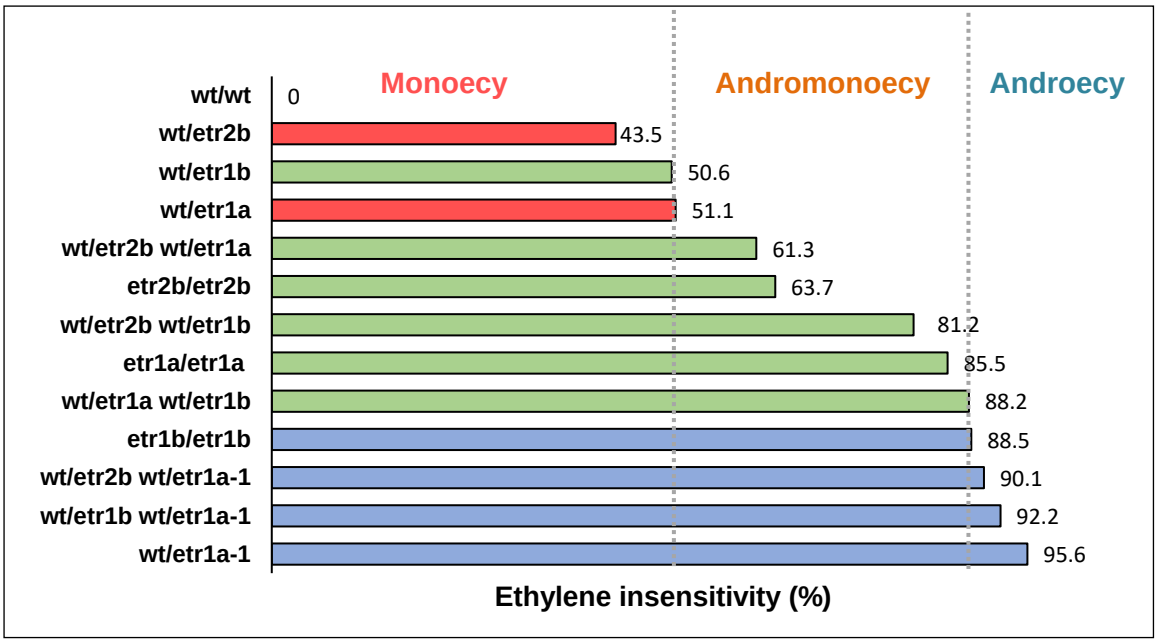


Figure 6. Correlation between the percentage of ethylene insensitivity in single and double mutants of *C. pepo* and the conferred sex morphotype: monoecy (red), andromonoecy (green) and androecy (blue).

The genetic network controlling sex determination in the *Cucurbitaceae* is beginning to be elucidated (Figure 7). At very early stages of flower development, the floral meristem possesses a default hermaphrodite-development pattern. However, whether a male or a female flower will develop is very quickly determined. Female flower development is achieved by two ethylene-regulated mechanisms: i) the arrest of stamen development primordia, a function controlled by the female-specific gene *ACS2/7* (Boualem et al., 2008, 2009; Ji et al., 2016; Manzano et al., 2016; Martínez et al., 2014), ii) the promotion of carpel development, a function that is achieved by the genes *ACS11* and *ACO2* (Boualem et al., 2015; Chen et al., 2016). Our findings indicate that these two ethylene-regulated pathways converge at the ethylene receptor genes. The induced ethylene response eventually activates the development of carpels, but arrests the development of stamens (Figure 7).

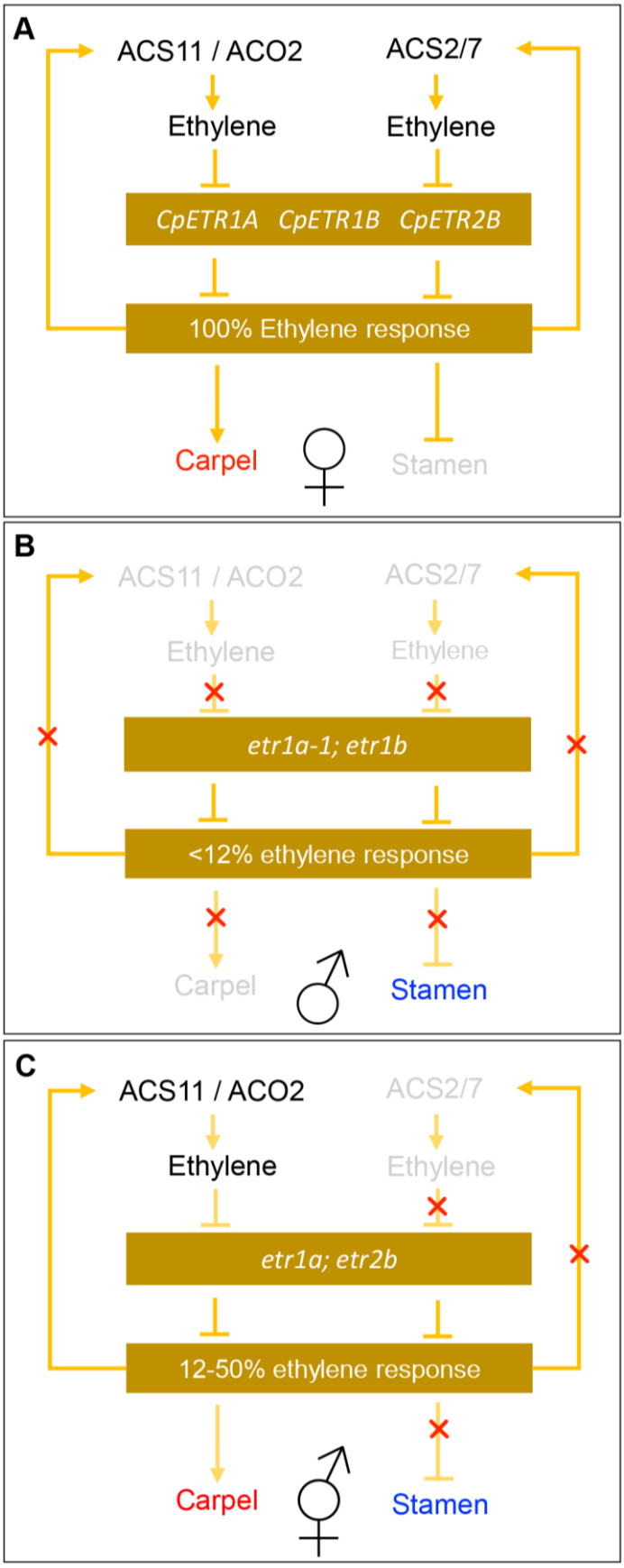


Figure 7. Model integrating ethylene biosynthesis and perception genes in the control of sex determination in *C. pepo*. (**A**) The genes *ACS11* and *ACS2/7* are involved in the biosynthesis of ethylene in the floral meristems that will be determined as female flowers. This ethylene is perceived by the receptors CpETR1A, CpETR1B and CpETR2, suppressing the negative regulation of ethylene receptors over the ethylene response. The 100% activated response to ethylene promotes the development of carpels and arrest the development of stamens, which results in the development of a female flower and in monoecy. The ethylene biosynthesis genes *ACS11* and *ACS2/7* are positively auto-regulated by ethylene response in *C. pepo*. (**B**) The *etr* mutations that reduce ethylene response below 12%, are able to disrupt the promotion of carpel and the arrest of stamen development, which results in the development of a male flower and to androecy. The lack of ethylene response also suppress the activation of ethylene biosynthesis genes *ACS11* and *ACS2/7* in the FM. (**C**) The *etr* mutations that reduce ethylene response between 12 and 50% are able to disrupt the arrest of stamen development, but the promotion of carpel is still occurring, which result in the development of an hermaphrodite flower and andromonoecy. In this case it is hypothesized that this reduced response to ethylene is able to suppress the activation of *ACS2/7* but not the activation of *ACS11* in the FM.

Results also indicate that the arrest of stamen development is more sensitive to a reduction in ethylene response than the promotion of carpel development, this indicates that stamens are more sensitive to ethylene than carpels, and therefore that carpels require a higher level of ethylene for a response to be evoked. Our gene expression data in the apical shoots of WT plants also support this conclusion, since the expression level of *CpACS11* and *CpACO2* was found to be much higher than that of *CpACS2*/7 (Figure 5).

On the other hand, in melon and cucumber it has been reported that male flower development requires the action of *WIP1*, a transcription factor that is responsible for the arrest of carpel development. Its disfunction in melon, cucumber and watermelon results in plants bearing only female flowers (gynoecy) (Boualem et al., 2008). In those species, the gene *WIP1* is known to be repressed by *ACS11* and *ACO2*, and that the lack of either of these allows the upregulation of *WIP1*. This arrests carpel development and represses gene *ACS2/7*, thus triggering the development of male flowers (Boualem et al., 2015; Tao et al., 2018). We found that the reduced ethylene response in the *etr1a-1* and *etr2b* androecious mutants prevents the upregulation of *CpACS11*, *CpACO2A* and *CpACS2/7A* during female flowering transition; however, we found no upregulation of *CpWIP1B*, the only *WIP1* ortholog that was found to be expressed. This suggests that in the absence of ethylene response, there must be an alternative pathway that activates male flower development, independently of *WIP1*.

Ethylene perception and signaling also control the time at which the production of male or female flowers takes place. In the single and double ethylene-insensitive mutants, female flower transition, i.e. the number of male flowers occurring before the production of the first female flower, was correlated with ethylene sensitivity. In monoecious plants, the transition occurred earlier than in andromonoecius mutants, and in the androecious mutants it was completely blocked (Table 3). This transition also resulted in the induction of the ethylene biosynthesis genes which are involved in sex determination (*CpACS11*, *CpACO2A* and *CpACS2/7A*), as observed in the apical shoots of WT plants upon transition to female flowering (Figure 5). In the androecious mutants, female flowering transition is blocked by either the suppression of ethylene response or by the prevention of the upregulation of *CpACSS11*, *CpACO2A* and *CpACS2/7A* in the apical shoots. This implies that these *ACS* genes are positively autoregulated by the action of ethylene (Figure 7).

6. General conclusions

1. The ethyl methanesulfonate (EMS) mutant *Cucurbita pepo* platform developed contains 3,751 independent lines with an average mutation frequency of 1/111 kb. The mutations are regularly distributed along the complete genome and consist of more than 75% of C>T and G>A transitions, which are the canonical mutations generated by EMS.

2. The predicted impact of WGS-detected mutations on gene functioning (0.4% moderate; 0.2% high) and the high percentage of phenotypic variation observed in the new squash-mutant platform (10.82% in seedlings, and 28% in adult plants), demonstrate the usefulness of the collection in high-throughput mutation research, regarding both gene function analysis and plant breeding.

3. The high-throughput screening of the whole mutant library by using the triple response to ethylene has yielded a total of four ethylene-insensitive mutants, identified as *ein1* (*etr1b*), *ein2* (*etr1a*), *ein3* (e*tr2b*) and *EIN4* (*etr1a-1*). The mutations *etr1a*, *etr1b* and *etr2b* are semidominant, while *etr1a-1* is dominant.

4. The mutations *etr1a* and *etr2b* modified sex determination mechanisms, promoting conversion of female flowers into bisexual or hermaphrodite ones, and monoecy into andromonoecy, thereby delaying the transition to female flowering and reducing the number of pistillate flowers per plant.

5. The mutations *etr1b* and *etr1a-1* block female flowering transition, promoting the conversion of monoecy into androecy.

6. The four *etr* mutations affect the growth rate and maturity of petals and carpels of pistillate flowers, either under homozygous or heterozygous conditions, thus lengthening the time required for flowers to reach anthesis. They also stimulate the growth rate of ovaries, and the parthenocarpic development of fruits.

7. Whole-genome re-sequencing of WT and mutant bulks allowed the identification of the causal mutations of the mutant phenotypes as being four missense mutations that affect the ethylene receptors genes *CpETR1A, CpETR1B* and *CpETR2B*, converting the receptors into a constitutive signaling-on state that repress the response to ethylene. The *etr1a*, *etr1a-1* and *etr1b* mutations disrupt the transmembrane ethylene binding domain

of CpETR1A and CpETR1B, while *etr2b* affects the coiled-coil domain between the GAF and histidine-kinase domains of CpETR2B, probably disrupting the transduction of ethylene signal.

8. The phenotypes of homozygous and heterozygous single- and double-mutant plants indicated that the ethylene receptor genes *CpETR1A*, *CpETR1B* and *CpETR2A* are not redundant, but cooperate in the control of the ethylene response and sex determination.

9. The level of ethylene insensitivity in single and double mutants is dependent upon the strength and the dosage of *etr* mutant alleles, and it determines the final sex phenotype of the plant. This ranges from monoecy in ethylene-sensitive WT plants, to andromonoecy in the partially ethylene-insensitive plants, and to androecy in the plants with highest level of insensitivity to ethylene.

10. The blocking of female flower development by the stronger *etr1a-1* and *etr1b* mutants is associated with a downregulation of *CpACS11*, *CpACO2* and *CpACS2*/7, three ethylene biosynthesis genes required for female flowering transition to take place in *C. pepo*.

7. Supplementary material

Supplementary Table 3.1 (3.3.4). Sequence and coverage statistics. Number of sequenced, unmapped and duplicate reads. Mean and median coverage and percent of genomic bases with a fold coverage higher than 5.

Sample	Seq. reads	Unmap reads	% Unmap	Dup. reads	% dup	Mean coverage	Median coverage	% Coverage >5x
L1.1	17,470,139	1,189,573	6.8	1,289,252	7.37	16.37	9	82.1
L1.2	17,472,279	1,100,148	6.29	1,272,279	7.28	16.38	10	82.6
L1.3	15,557,026	954,025	6.13	1,054,410	6.77	14.64	9	78.9
L1.4	17,434,357	1354914	7.77	1,278,591	7.33	16.33	9	82.2
L2.1	18,583,646	1253795	6.74	1,410,944	7.59	17.39	10	83.7
L2.2	18,245,737	1,055,990	5.78	1,373,955	7.53	17.08	10	82.9
L2.3	19,185,343	1,046,213	5.45	1,557,843	8.11	17.9	10	83.8
L2.4	13,983,971	870,083	6.22	903,141	6.45	13.18	8	73

Supplementary Table 3.2 (3.4.5). Genes affected by high and moderate impact mutations in squash L1 mutant family

Impact	Mutation type	Transcript ID	Description
High	Splice acceptor variant	Cp4.1LG05g14810.1	probable phosphatase 2C 52
		Cp4.1LG12g03450.1	YTH domain-containing family 3-like isoform X1
	Splice donor variant	Cp4.1LG03g18020.1	myb family transcription factor APL isoform X1
		Cp4.1LG04g06490.1	acetyl-coenzyme A chloroplastic glyoxysomal-like
		Cp4.1LG06g08460.1	transcription termination factor chloroplastic-like
		Cp4.1LG17g09630.1	myosin-6-like
		Cp4.1LG18g09120.1	
	Stop gained	Cp4.1LG03g06410.1	glyceraldehyde-3-phosphate dehydrogenase chloroplastic
		Cp4.1LG05g15120.1	multi -bridging factor 1a-like
		Cp4.1LG06g08230.1	probable LRR receptor-like serine threonine- kinase At4g26540
		Cp4.1LG07g03960.1	low-temperature-induced cysteinase
		Cp4.1LG08g08410.1	nuclear transcription factor Y subunit B-9
		Cp4.1LG09g00960.1	ADP-ribosylation factor-related 1
		Cp4.1LG09g04840.1	Protease 2
		Cp4.1LG13g00490.1	
		Cp4.1LG17g05510.1	aquaporin TIP2-1
Moderate	Missense variant	Cp4.1LG01g00800.1	9-cis-epoxycarotenoid dioxygenase chloroplastic-like
		Cp4.1LG01g01400.1	
		Cp4.1LG01g05320.1	probable phosphatase 2C 25
		Cp4.1LG01g06760.1	CHAPERONE-LIKE PROTEIN OF chloroplastic isoform X1
		Cp4.1LG01g08050.1	DNA polymerase I chloroplastic mitochondrial
		Cp4.1LG01g11140.1	LIM domain-containing WLIM2b
		Cp4.1LG01g14610.1	cysteine ase 15A
		Cp4.1LG01g14870.1	Myb_DNA-binding domain-containing
		Cp4.1LG01g18540.1	O-fucosyltransferase family
		Cp4.1LG01g18720.1	
		Cp4.1LG01g19610.1	DNA-directed RNA polymerase II non-catalytic subunit
		Cp4.1LG01g19940.1	
		Cp4.1LG01g20300.1	biotin-- ligase 2-like
		Cp4.1LG01g24610.1	ATP binding
		Cp4.1LG01g24630.1	
		Cp4.1LG02g03480.1	high affinity nitrate transporter
		Cp4.1LG02g05950.1	solute carrier family 25 member 44-like
		Cp4.1LG02g07960.1	cytokinin riboside 5 -monophosphate phosphoribohydrolase LOG1-like
		Cp4.1LG02g08980.1	Eukaryotic translation initiation factor 5B
		Cp4.1LG02g10530.1	diacylglycerol kinase 1
		Cp4.1LG02g10940.1	SCY1 2
		Cp4.1LG02g14220.1	phospholipase D alpha 4
		Cp4.1LG02g14230.1	
		Cp4.1LG03g00260.1	40S ribosomal S5
		Cp4.1LG03g01280.1	cytochrome P450 76A1
		Cp4.1LG03g02170.1	ABC transporter G family member 20-like
		Cp4.1LG03g03080.1	TIC chloroplastic
		Cp4.1LG03g06430.1	transportin-3 isoform X1
		Cp4.1LG03g06550.1	Phosphoglycerate mutase family
		Cp4.1LG03g07640.1	NHL domain-containing
		Cp4.1LG03g10330.1	CBL-interacting serine threonine- kinase 4-like
		Cp4.1LG03g11500.1	probable phosphatase 2C 40
		Cp4.1LG03g12080.1	Chromatin structure-remodeling complex BSH
		Cp4.1LG03g13690.1	LURP-one-related 5
		Cp4.1LG03g13870.1	boron transporter 4-like
		Cp4.1LG03g14650.1	probable GTP diphosphokinase chloroplastic
		Cp4.1LG03g15560.1	serine threonine- kinase D6PK
		Cp4.1LG03g17480.1	protoheme IX mitochondrial
		Cp4.1LG04g00600.1	serine--tRNA ligase
		Cp4.1LG04g00740.1	stress regulated
		Cp4.1LG04g00810.1	coatomer subunit gamma-2
		Cp4.1LG04g01140.1	probable transcription factor 21
		Cp4.1LG04g01280.1	probable pectate lyase P59
		Cp4.1LG04g06420.1	
		Cp4.1LG04g07180.1	GATA transcription factor 24-like isoform X1
		Cp4.1LG04g08210.1	myb family transcription factor EFM-like
		Cp4.1LG04g08360.1	cytochrome c biogenesis chloroplastic
		Cp4.1LG04g10480.1	phosphoinositide 3-kinase regulatory subunit 4
		Cp4.1LG04g10830.1	esterase lipase thioesterase family
		Cp4.1LG04g10900.1	CTR9 homolog
		Cp4.1LG04g11740.1	serrate RNA effector molecule
		Cp4.1LG04g14000.1	Uncharacterized RING finger
		Cp4.1LG04g15700.1	FKBP12-interacting of 37 kDa
		Cp4.1LG05g00860.1	subtilisin-like protease
		Cp4.1LG05g01590.1	receptor-like cytosolic serine threonine- kinase RBK2 isoform X1
		Cp4.1LG05g02890.1	superoxide dismutase [Cu-Zn] chloroplastic
		Cp4.1LG05g03210.1	respiratory burst oxidase homolog A
		Cp4.1LG05g04410.1	transcriptional activator DEMETER
		Cp4.1LG05g05210.1	casein kinase 1 HD16
		Cp4.1LG05g07160.1	Rubredoxin family

Moderate	Missense variant	Cp4.1LG05g08360.1	probable inactive ATP-dependent zinc metalloprotease FTSHI chloroplastic
		Cp4.1LG05g08630.1	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
		Cp4.1LG05g10760.1	
		Cp4.1LG05g12160.1	PREDICTED: uncharacterized protein LOC103498397 isoform X1
		Cp4.1LG05g12990.1	PREDICTED: uncharacterized protein LOC103497509
		Cp4.1LG05g13470.1	probable serine threonine- kinase At4g35230
		Cp4.1LG05g13660.1	
		Cp4.1LG05g14060.1	heterogeneous nuclear ribonucleo 1
		Cp4.1LG05g14450.1	syntaxin-132-like isoform X1
		Cp4.1LG05g14480.1	formin 3
		Cp4.1LG05g14830.1	external alternative NAD(P)H-ubiquinone oxidoreductase mitochondrial
		Cp4.1LG05t00050.1	
		Cp4.1LG05t01770.1	
		Cp4.1LG06g01480.1	Peptidase_S9 domain-containing
		Cp4.1LG06g02180.1	probable xyloglucan glycosyltransferase 12
		Cp4.1LG06g03720.1	
		Cp4.1LG06g04460.1	L-arabinokinase-like isoform X1
		Cp4.1LG06g04950.1	
		Cp4.1LG06g04960.1	
		Cp4.1LG06g05510.1	probable phosphoinositide phosphatase SAC9
		Cp4.1LG06g08320.1	pentatricopeptide repeat-containing At5g55840 isoform X1
		Cp4.1LG06g09150.1	
		Cp4.1LG07g01040.1	transcription factor TGA2-like
		Cp4.1LG07g01310.1	probable mitochondrial adenine nucleotide transporter BTL3
		Cp4.1LG07g02790.1	transcription factor HY5
		Cp4.1LG07g04920.1	subtilisin-like protease
		Cp4.1LG07g05070.1	
		Cp4.1LG07g05310.1	
		Cp4.1LG07g07740.1	U-box domain-containing 35 isoform X1
		Cp4.1LG07g07800.1	ethylene receptor
		Cp4.1LG07g08080.1	transcription elongation factor SPT6
		Cp4.1LG07g08090.1	histone-lysine N-methyltransferase SUVR4
		Cp4.1LG07g09900.1	myb family transcription factor EFM
		Cp4.1LG07g10790.1	auxin response factor 19
		Cp4.1LG07g11560.1	nuclear pore complex NUP205
		Cp4.1LG08g01320.1	U-box domain-containing 33
		Cp4.1LG08g02170.1	brefeldin A-inhibited guanine nucleotide-exchange 5
		Cp4.1LG08g04100.1	non-specific lipid-transfer At2g13820
		Cp4.1LG08g06260.1	BAG-associated GRAM 1
		Cp4.1LG08g06810.1	beta-glucosidase 3B-like
		Cp4.1LG08g07950.1	peroxisome biogenesis 5
		Cp4.1LG08g08160.1	
		Cp4.1LG08g08220.1	caffeic acid 3-O-methyltransferase-like
		Cp4.1LG08g11620.1	myb-related A-like
		Cp4.1LG09g00200.1	histone-lysine N-methyltransferase ATX4-like
		Cp4.1LG09g02870.1	Histone-lysine N-methyltransferase isoform 1
		Cp4.1LG09g06570.1	vesicle-fusing ATPase
		Cp4.1LG09g07160.1	Tubulin beta-1 chain
		Cp4.1LG09g07300.1	GPI ethanolamine phosphate transferase 3
		Cp4.1LG09g11580.1	LONGIFOLIA 1
		Cp4.1LG10g01950.1	DEAD-box ATP-dependent RNA helicase 16
		Cp4.1LG10g04010.1	probable UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase SPINDLY
		Cp4.1LG10g05900.1	probable 9-cis-epoxycarotenoid dioxygenase chloroplastic
		Cp4.1LG10g06660.1	
		Cp4.1LG10g06760.1	copper-transporting ATPase RAN1
		Cp4.1LG10g08880.1	flocculation FLO11
		Cp4.1LG10g09210.1	sphingosine kinase 1 isoform X1
		Cp4.1LG10g10030.1	E3 ubiquitin- ligase UPL5
		Cp4.1LG10g10050.1	rho GDP-dissociation inhibitor 1
		Cp4.1LG10g10690.1	
		Cp4.1LG10g11600.1	40S ribosomal S2-4
		Cp4.1LG11g00600.1	probable alpha-galactosidase B
		Cp4.1LG11g04620.1	E3 ubiquitin- ligase UPL2-like
		Cp4.1LG11g04940.1	rho GTPase-activating REN1
		Cp4.1LG11g06140.1	FPL domain-containing
		Cp4.1LG11g06530.1	beta-hexosaminidase 3
		Cp4.1LG11g06580.1	
		Cp4.1LG11g07430.1	cyclin delta-1 family
		Cp4.1LG11g09780.1	GDP-mannose 3
		Cp4.1LG11g10050.1	myosin-binding 1-like
		Cp4.1LG11g10080.1	
		Cp4.1LG12g00190.1	NRT1 PTR FAMILY -like
		Cp4.1LG12g05480.1	CW-type Zinc isoform 1
		Cp4.1LG12g06240.1	
		Cp4.1LG12g06660.1	structural maintenance of chromosomes 5
		Cp4.1LG12g07930.1	
		Cp4.1LG12g08740.1	inorganic pyrophosphatase 2-like
		Cp4.1LG13g00070.1	
		Cp4.1LG13g02200.1	
		Cp4.1LG13g03850.1	GIGANTEA-like

Moderate	Missense variants	Cp4.1LG13g04830.1	DIS3-like exonuclease 2
		Cp4.1LG13g05860.1	alpha-mannosidase isoform X1
		Cp4.1LG13g07940.1	6-phosphogluconate decarboxylating 3
		Cp4.1LG13g08100.1	DNA-directed RNA polymerase chloroplastic mitochondrial
		Cp4.1LG13g10680.1	
		Cp4.1LG13g11170.1	protoporphyrinogen chloroplastic mitochondrial isoform X1
		Cp4.1LG14g02860.1	
		Cp4.1LG14g04120.1	pinin
		Cp4.1LG14g04770.1	zinc transport
		Cp4.1LG14g05560.1	embryo yellow
		Cp4.1LG14g05840.1	receptor-like serine threonine- kinase At4g25390
		Cp4.1LG14g05910.1	
		Cp4.1LG14g06120.1	
		Cp4.1LG14g06300.1	
		Cp4.1LG14g06470.1	dnaJ ERDJ2A-like
		Cp4.1LG14g06560.1	
		Cp4.1LG14g09530.1	scarecrow 22
		Cp4.1LG15g00880.1	CHROMATIN REMODELING 8 isoform X1
		Cp4.1LG15g01750.1	chloroplastic
		Cp4.1LG15g04110.1	probable S-acyltransferase 23 isoform X1
		Cp4.1LG15g06990.1	
		Cp4.1LG15g07390.1	
		Cp4.1LG15g08420.1	
		Cp4.1LG15g09050.1	nuclear-pore anchor
		Cp4.1LG16g00510.1	
		Cp4.1LG16g01290.1	
		Cp4.1LG16g06480.1	
		Cp4.1LG16g07930.1	serine threonine- kinase D6PKL1-like
		Cp4.1LG17g02130.1	NRT1 PTR FAMILY -like
		Cp4.1LG17g03160.1	vesicle-associated membrane 722
		Cp4.1LG17g03710.1	
		Cp4.1LG17g05320.1	SAWADEE HOMEODOMAIN HOMOLOG 1
		Cp4.1LG17g05360.1	ankyrin repeat-containing At5g02620-like
		Cp4.1LG17g05640.1	
		Cp4.1LG17g06840.1	CW-type Zinc isoform 1
		Cp4.1LG17g07150.1	
		Cp4.1LG17g08430.1	
		Cp4.1LG17g08930.1	D-glycerate 3- chloroplastic
		Cp4.1LG17g10040.1	
		Cp4.1LG18g00520.1	phosphoacetylglucosamine mutase
		Cp4.1LG18g02640.1	YTH domain-containing family 1 isoform X1
		Cp4.1LG18g05950.1	PTI1-like tyrosine- kinase At3g15890
		Cp4.1LG18g06420.1	tRNA-specific adenosine deaminase 2 isoform X1
		Cp4.1LG18g07140.1	
		Cp4.1LG18g07200.1	tip elongation aberrant 1-like
		Cp4.1LG18g08060.1	NRT1 PTR FAMILY -like
		Cp4.1LG19g00610.1	
		Cp4.1LG19g02570.1	pyruvate decarboxylase 1-like
		Cp4.1LG19g03040.1	DEFECTIVE IN EXINE FORMATION 1
		Cp4.1LG19g05460.1	chloroplastic-like
		Cp4.1LG19g07470.1	RNA polymerase II C-terminal domain phosphatase-like 1
		Cp4.1LG19g10100.1	kinesin KCA2
		Cp4.1LG19g10380.1	
		Cp4.1LG19g10580.1	
		Cp4.1LG19g11280.1	O-Glycosyl hydrolases family 17 isoform
		Cp4.1LG20g00070.1	transcriptional activator Myb-like
		Cp4.1LG20g00310.1	
		Cp4.1LG20g00480.1	sucrose synthase 7-like
		Cp4.1LG20g01050.1	phospholipid-transporting ATPase 9
		Cp4.1LG20g01690.1	
		Cp4.1LG20g04530.1	
		Cp4.1LG20g04770.1	
		Cp4.1LG20g05600.1	30-kDa cleavage and polyadenylation specificity factor 30
		Cp4.1LG20g07500.1	

Supplementary Table 3.3 (3.4.5). Genes affected by high and moderate impact mutations in squash L2 mutant family

Impact	Mutation type	Transcript ID	Description
High	Splice acceptor variant	Cp4.1LG12g11380.1	auxin-induced 22D-like
	Stop gained	Cp4.1LG08g11700.1	E3 ubiquitin- ligase PRT6
		Cp4.1LG17g01540.1	probable RNA helicase SDE3
Moderate	Missense variant	Cp4.1LG01g01650.1	
		Cp4.1LG01g02780.1	cyclin d
		Cp4.1LG01g03030.1	root phototropism 3
		Cp4.1LG01g03490.1	lipase YOR059C
		Cp4.1LG01g06500.1	zinc finger 4-like
		Cp4.1LG01g07560.1	1-phosphatidylinositol-3-phosphate 5-kinase FAB1C
		Cp4.1LG01g10540.1	
		Cp4.1LG01g13880.1	Hemerythrin domain-containing zf-CHY domain-containing zf-RING_2 domain-containing
		Cp4.1LG01g25240.1	
		Cp4.1LG02g02290.1	callose synthase 8
		Cp4.1LG02g04490.1	DUF490 domain-containing
		Cp4.1LG02g06460.1	RNA-binding 1-like
		Cp4.1LG02g06770.1	heat stress transcription factor B-4
		Cp4.1LG02g14160.1	
		Cp4.1LG02g14740.1	EPIDERMAL PATTERNING FACTOR 8
		Cp4.1LG02g16030.1	
		Cp4.1LG03g04070.1	
		Cp4.1LG03g07250.1	[-PII] uridylyltransferase
		Cp4.1LG03g10030.1	ethylene receptor 2
		Cp4.1LG03g16030.1	CBL-interacting serine threonine- kinase 9
		Cp4.1LG03g16830.1	DEAD-box ATP-dependent RNA helicase 8
		Cp4.1LG03t02430.1	
		Cp4.1LG04g04810.1	serine threonine- kinase STY46-like
		Cp4.1LG04g09990.1	
		Cp4.1LG05g02460.1	probable aminotransferase ACS10
		Cp4.1LG05g03820.1	
		Cp4.1LG05g04020.1	
		Cp4.1LG05g11020.1	
		Cp4.1LG05g13070.1	eukaryotic translation initiation factor 4B3-like
		Cp4.1LG05g14200.1	
		Cp4.1LG05t01310.1	
		Cp4.1LG06g04820.1	HBS1 isoform X1
		Cp4.1LG06g08100.1	carbamoyl-phosphate synthase large chloroplastic
		Cp4.1LG06g08270.1	E3 ubiquitin- ligase At4g11680-like
		Cp4.1LG06t01480.1	
		Cp4.1LG06t02530.1	
		Cp4.1LG07g05040.1	arginine N-methyltransferase 2
		Cp4.1LG07g07510.1	
		Cp4.1LG07g11560.1	nuclear pore complex NUP205
		Cp4.1LG08g03280.1	
		Cp4.1LG08g07510.1	probable trehalose-phosphate phosphatase F
		Cp4.1LG08g09050.1	probable choline kinase 2
		Cp4.1LG08g10860.1	ABC transporter B family member 1
		Cp4.1LG09g10100.1	tryptophan aminotransferase-related 2
		Cp4.1LG10g01680.1	
		Cp4.1LG10g06380.1	dihydroceramide fatty acyl 2-hydroxylase FAH1
		Cp4.1LG10g10710.1	4-alpha-glucanotransferase DPE2
		Cp4.1LG11g04810.1	
		Cp4.1LG11g10050.1	myosin-binding 1-like
		Cp4.1LG12g04910.1	importin beta-like SAD2
		Cp4.1LG12g10010.1	
		Cp4.1LG13g05110.1	probable tRNA modification GTPase
		Cp4.1LG13g05600.1	lysine-specific demethylase JMJ25-like
		Cp4.1LG13g07670.1	E3 ubiquitin- ligase SHPRH isoform X1
		Cp4.1LG13g08860.1	WAT1-related At3g28050-like

Moderate	Missense variant	Cp4.1LG13g10730.1	
		Cp4.1LG14g07170.1	small G signaling modulator 2-like isoform X3
		Cp4.1LG14g07730.1	CWC15 homolog
		Cp4.1LG14g08580.1	serine threonine- kinase rio2
		Cp4.1LG15g01720.1	
		Cp4.1LG15g04030.1	E3 ubiquitin- ligase RING1-like
		Cp4.1LG15g06300.1	pectinesterase-like
		Cp4.1LG15g07330.1	fatty acid hydroperoxide chloroplastic
		Cp4.1LG15g07500.1	
		Cp4.1LG15g08310.1	
		Cp4.1LG17g00420.1	lipase class 3 family
		Cp4.1LG17g00810.1	
		Cp4.1LG17g01400.1	mavicyanin-like
		Cp4.1LG17g01920.1	
		Cp4.1LG17g10520.1	serine threonine- kinase ATG1
		Cp4.1LG18g02790.1	
		Cp4.1LG18g09450.1	
		Cp4.1LG19g06270.1	
		Cp4.1LG19g07550.1	G-type lectin S-receptor-like serine threonine- kinase B120 isoform X1

Supplementary Table 3.4 (3.4.4). Filtering steps of the original sequencing data of L1 and L2 mutant families to discard false positive mutations induced by EMS.

	L1		**L2**		**L1 + L2**	
	Base coverage (% Ref. genome)	**No. SNVs**	**Base coverage (% Ref. genome)**	**No. SNVs**	**Shared bases (L1 and L2)**	**Shared SNVs (L1 and L2)**
BQ>20	198,385,685 76.10%		198451381 76.10%			
Filtering out bases with coverage lower than 8 and higher than 100	174752501 67%	190,333	176805948 67.80%	183,167	162504354 (62.3%)	
Filtering regions with excess of variants: more than 1 SNV/kb	157579060 60.40%	6,140	160330390 61.50%	5,257	144477323 (55.4%)	4,031
Filtering out common SVNs in shared bases between L1 and L2	144477323 55.40%	2,109	144477323 55.40%	1,226		
Homozygous SVNs		405		367		
Heterozygous SVNs: **EMS mutations**		**1704 (11.8/Mb)**		**859 (6/Mb)**		

Supplementary Table 4.1 (4.3.3, 4.4.4). Sequence and coverage statistics. Number of sequences, mapped reads, mean coverage and percentage of genomic bases with a fold coverage higher than 5, and number of SNPs per sample based on quality and depth. Filtering steps used to search for the putative causal mutation of *etr1a* and *etr2b* mutant phenotypes.

	etr1a family		***etr2b family***	
	WT-like bulk	**Mutant-like bulk**	**WT-like bulk**	**Mutant-like bulk**
Phenotype	Monoecious	Andromonoecious	Monoecious	Andromonoecious
***C. pepo* reference genome**	C.pepo_v3.2	C.pepo_v3.2	C.pepo_v3.2	C.pepo_v3.2
Nb of paired reads	34,942,418	32,991,383	36,829,383	33,169,314
% mapped reads	93.5	93.1	93.7	94.2
Mean coverage	16.4	15.5	17.2	15.5
% Coverage >5X	82.4	80.6	83.3	78.4
No. of SNPs (GQ>20)	43,886	42,611	44,076	40,315
EMS mutations (CT>GA)	15,045	14,938	15,134	15,253
Family specific mutations	1,074	967	1,163	1,282
Putative mutations (DP>4, 0.2<AF>0.8)		2		1

Supplementary Table 4.2 (4.3.4). Primers and TaqMan probes used for genotyping *etr1a* and *etr2b* mutations.

etr1a family	**Sequence**	**Tm (°C)**
Forward	GTTCAGTTTGGTGCTTTCAT	54.3
Reverse	ACAAGCATAAGCGCAGTC	53.8
Probe C	FAM — TAACCGCTGTGGTATCGTGTGC — BHQ1	64.2
Probe T	HEX— TAACCGCTGTGGTATCGTGTGC —BHQ1	64.2
***etr2b* family**		
Forward	TAGCTTGGCTTGCCATCA	53.8
Reverse	TCTTGGAGTAACCAGGAAC	55
Probe C	FAM —AGCTGGGACTCCTCCAGAAG— BHQ1	62.5
Probe T	HEX—AGCTGGGACTTCTCCAGAAG—BHQ1	60.5

Supplementary Table 4.3 (4.3.5). Primers for qRT-PCR analysis.

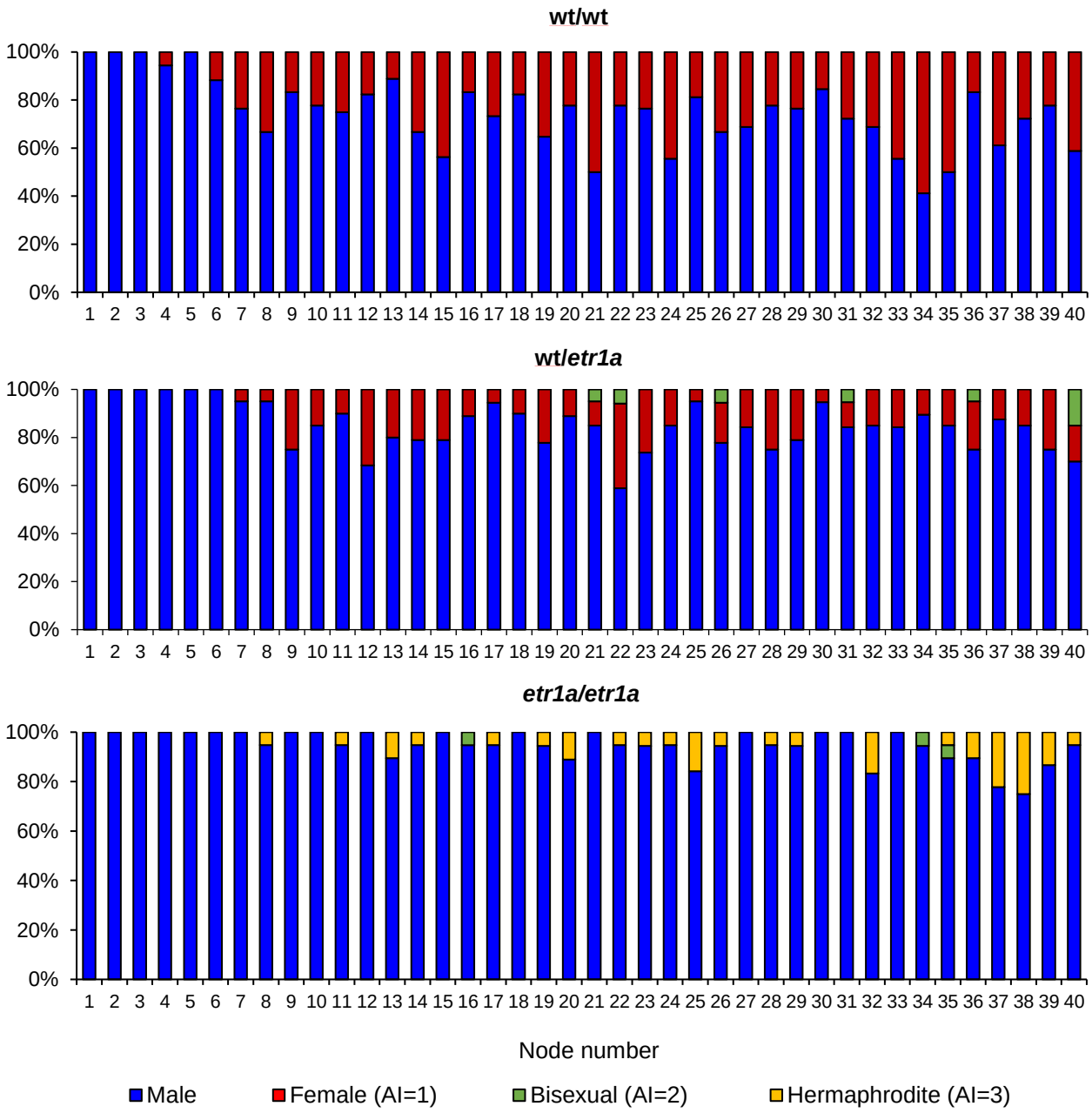
Gene	Primer name	Sequence
CpEf-1α	EF-1α-F	CGTCAAGAAGAAATAAGCCA
	EF-1α-R	CTACTACGAGAGAGAGAGCCG
CpETR1A	CpETR1.1-F	AAAGGAGAGCTGCCTGAGAGTC
	CpETR1.1-R	CACGACGCTCTATAAGTTCCGA
CpETR1B	CpETR1.2-F	GAGCGTCGGGTTCTATTCGA
	CpETR1.2-R	AACCTGGGATATGCCTTGTATGTAC
CpETR2B	CpETR2.1-F	TGAAGACTGGGAGAGATGC
	CpETR2.1-R1	CGGTTCCATTCATCCAAT
CpETR2A	CpETR2.2-F1	GATCAAAAAATGGCTTCTGC
	CpETR2.2-R	GAGACCTACTCGACTCGATAGA

Supplementary Table 4.4 (4.3.6, 4.4.5). Proteins used to perform the phylogenetic analysis.

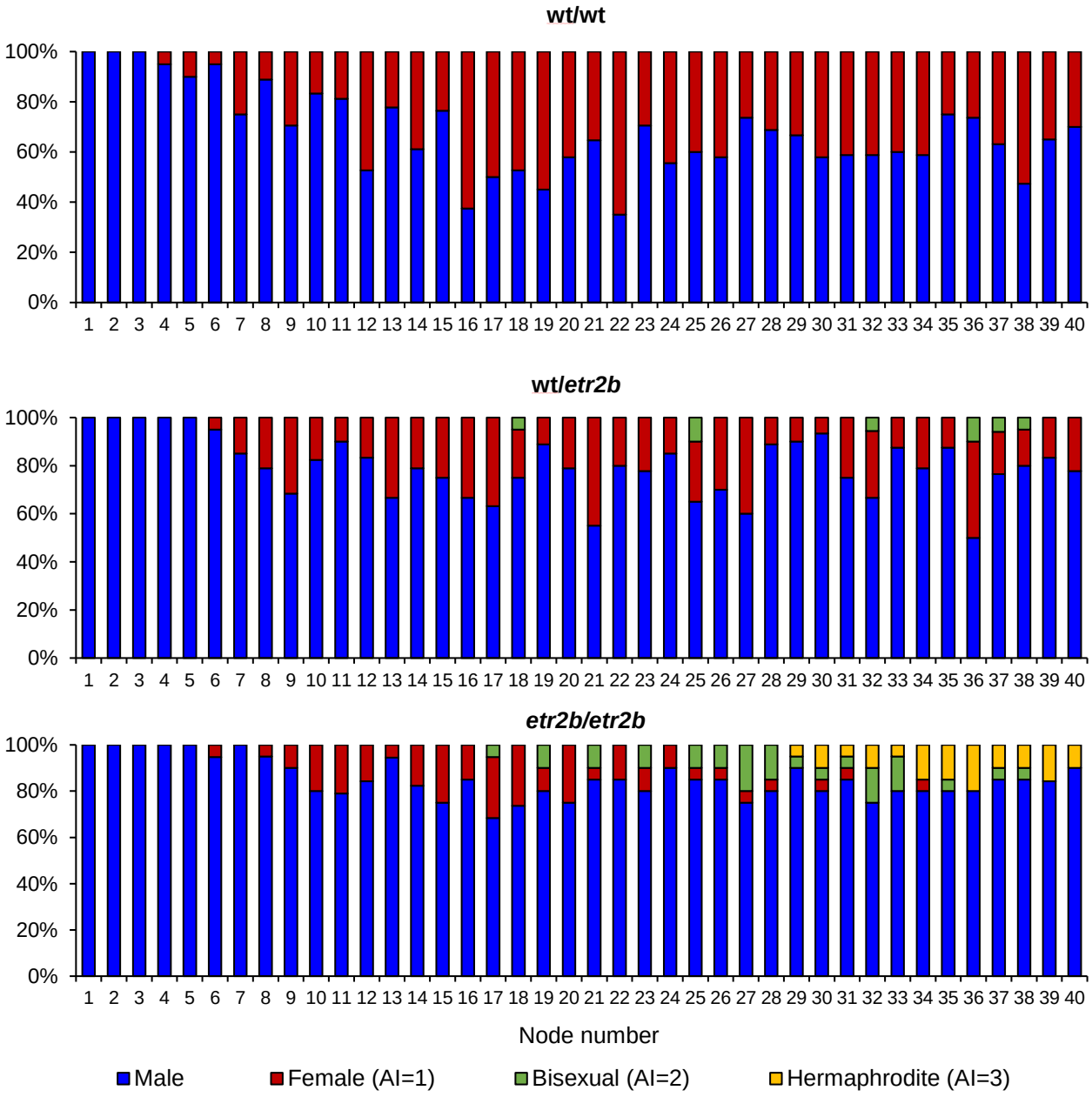
Species	Transcript Name	Abbreviation
Arabidopsis thaliana	AT1G66340	***AtETR1***
	AT3G04580	***AtEIN4***
	AT2G40940	***AtERS1***
	AT1G04310	***AtERS2***
	AT3G23150	***AtETR2***
Cucumis sativus	Cucsa.178860	***CsETR1***
	Cucsa.255140	***CsETR2***
	Cucsa.205330	***CsERS1***
Cucumis melo	MELO3C003906	***CmETR1***
	MELO3C006451	***CmETR2***
	MELO3C015961	***CmERS1***
Citrullus lanatus	ClCG11G011170	***ClETR1***
	ClCG05G004500	***ClETR2***
	ClCG09G005500	***ClERS1***
Cucurbita pepo	Cp4.1LG07g07800	***CpETR1A***
	Cp4.1LG11g06650	***CpETR1B***
	Cp4.1LG03g10030	***CpETR2B***
	Cp4.1LG08g03550	***CpETR2A***
	Cp4.1LG12g05940	***CpERS1B***
	Cp4.1LG17g07380	***CpERS1A***
Cucurbita moschata	CmoCh12G009400	***CmosETR1A***
	CmoCh05G007780	***CmosETR1B***
	CmoCh14G018320	***CmosETR2B***
	CmoCh06G013460	***CmosETR2A***
	CmoCh17G010370	***CmosERS1B***
	CmoCh08G004320	***CmosERS1A***
Cucurbita maxima	CmaCh12G009390	***CmaxETR1A***
	CmaCh05G007380	***CmaxETR1B***
	CmaCh14G017940	***CmaxETR2B***
	CmaCh00G005460	***CmaxETR2A***
	CmaCh17G010640	***CmaxERS1B***
	CmaCh08G004430	***CmaxERS1A***

Supplementary Table 4.5 (4.4.1). Segregation of ethylene sensitive, -intermediate and -insensitive plants in the offspring of backcrossed (BC_1 and BC_2) and selfed (BC_1S_1 and BC_2S_1) generations.

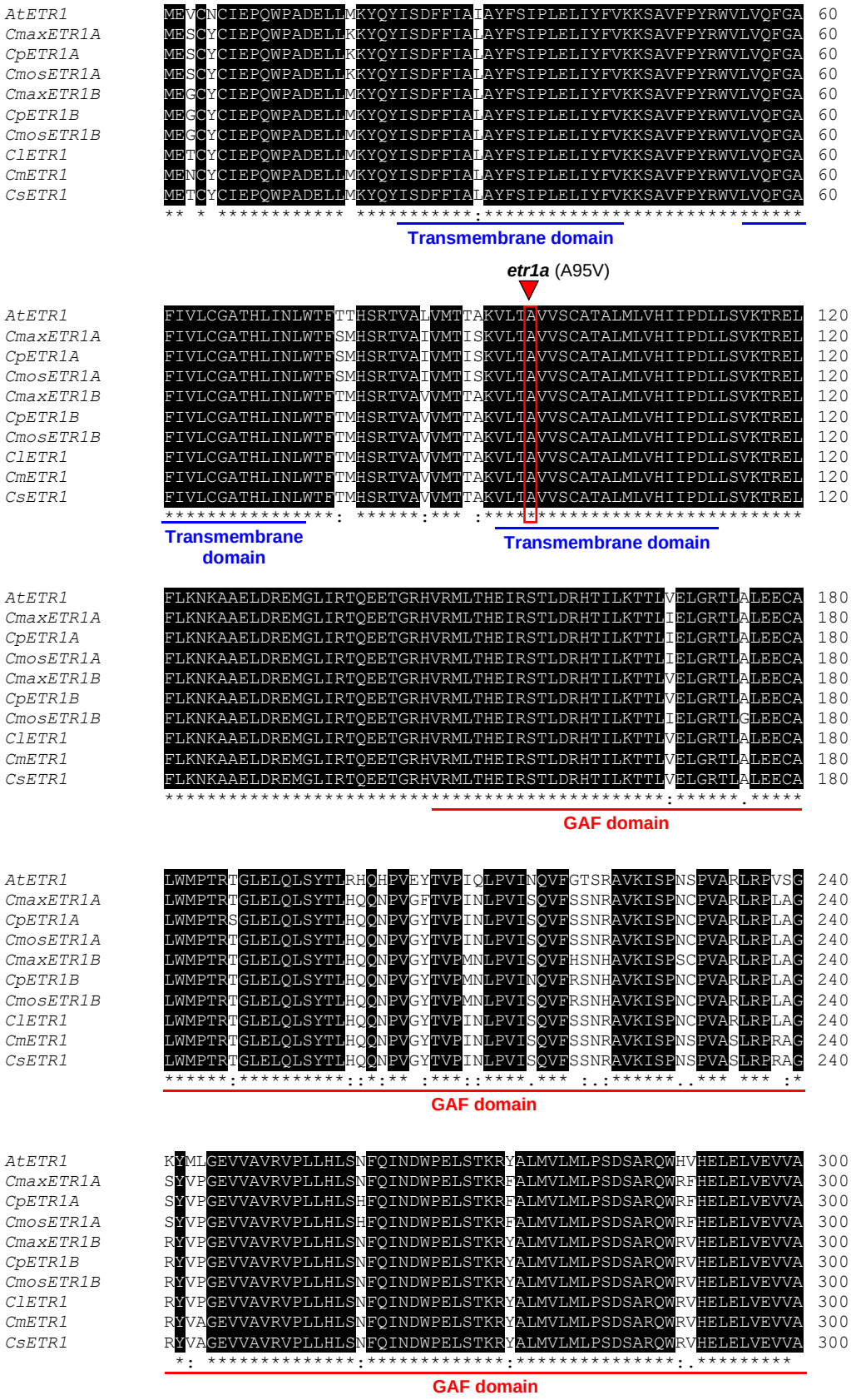
Mutant family	Generation	Number of plants			Expected segregation	χ^2	p-value
		Sensitive (WT)	Intermediate (wt/*etr*)	Insensitive (*etr*/*etr*)			
etr1a	BC_1	-	60	-	0:1:0	-	-
	BC_2	45	47	-	1:1:0	0.04	0.83
	BC_1S_1	132	278	141	1:2:1	0.34	0.84
	BC_2S_1	129	253	121	1:2:1	0.27	0.87
etr2b	BC_1	-	56	-	0:1:0		
	BC_2	55	50	-	1:1:0	0.24	0.63
	BC_1S_1	113	225	105	1:2:1	0.4	0.82
	BC_2S_1	121	252	127	1:2:1	0.18	0.92



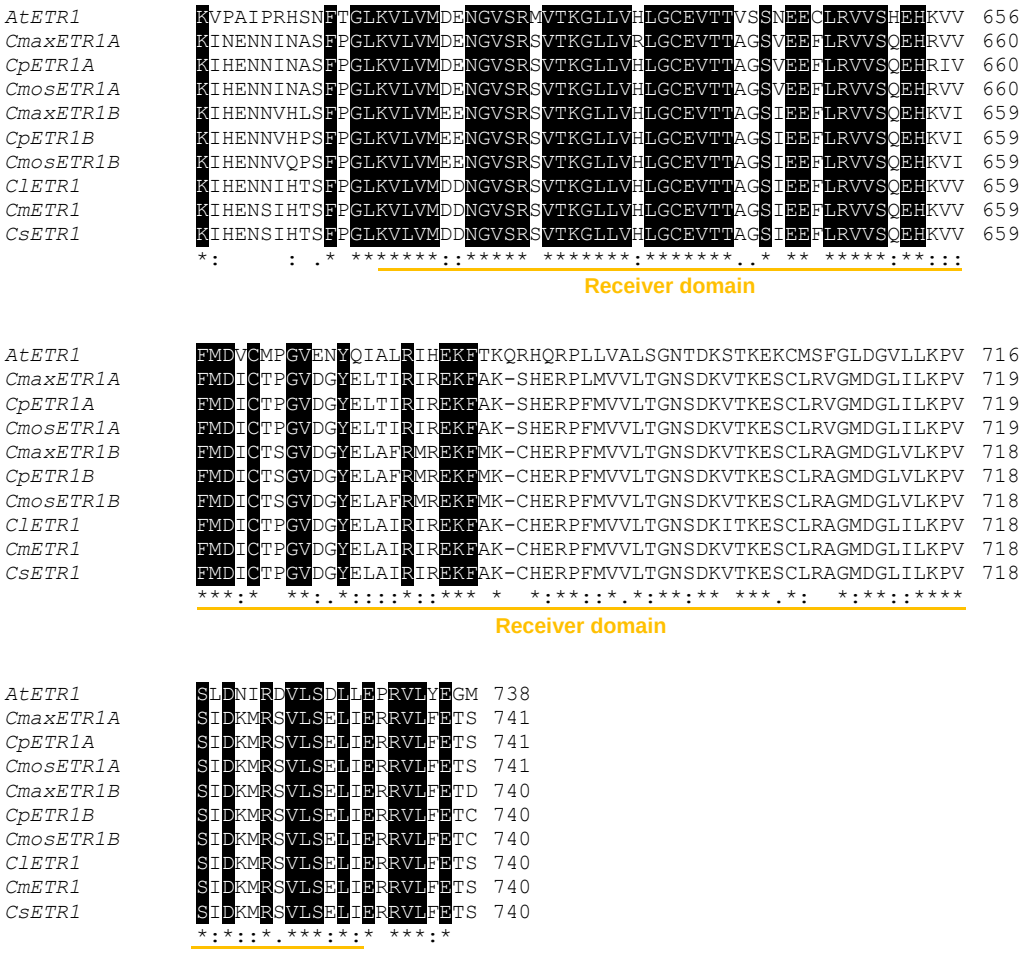
Supplementary Figure 4.1 (4.4.1). Distribution of staminate and pistillate flowers in the 40 first nodes of the plant. In each node, blue, red, green and yellow bars represent the percentages of male, female, bisexual and hermaphrodite flowers in the total number of plants analyzed in *etr1a* and *etr2b* mutant families (30 plants for each genotype).



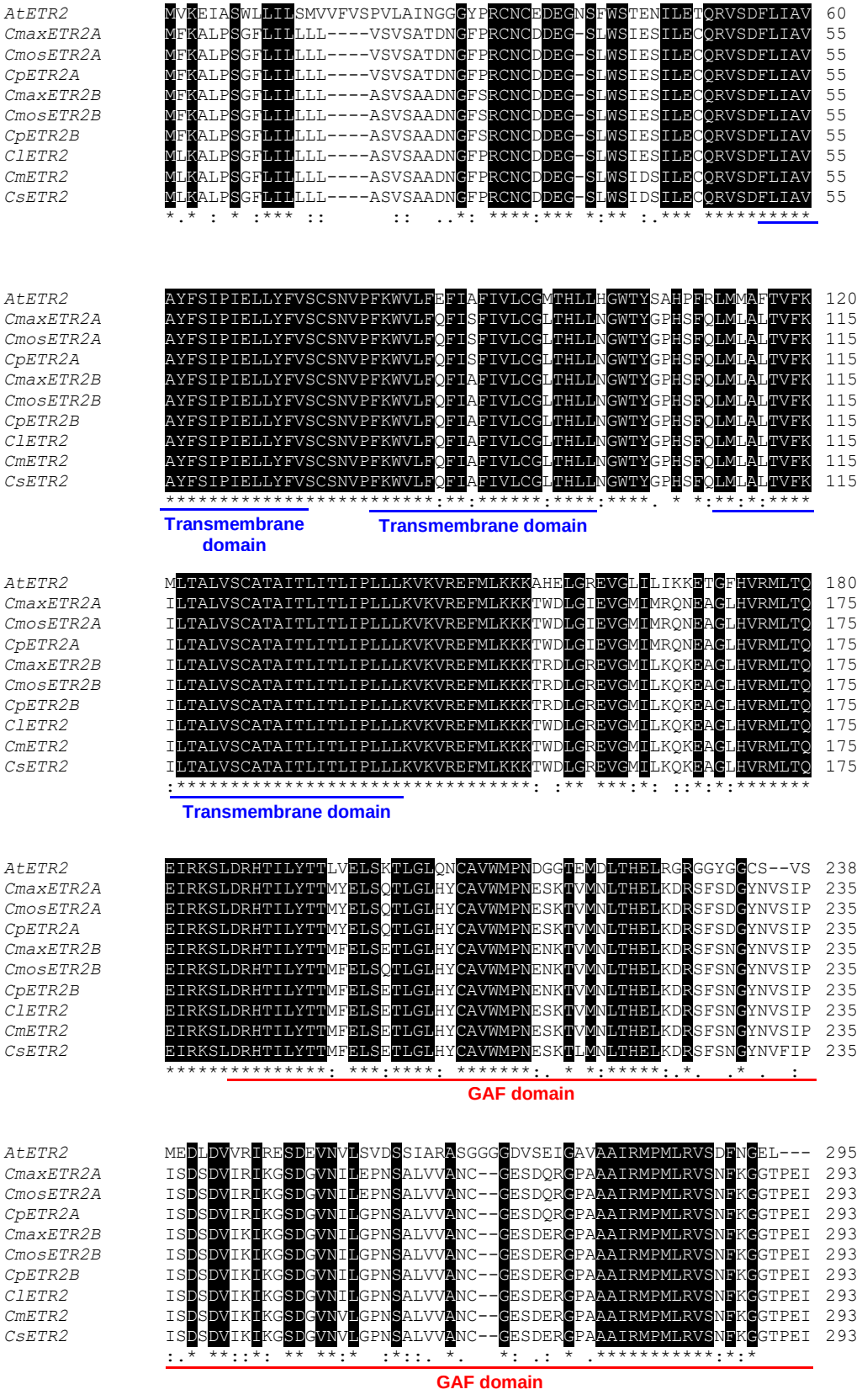
Supplementary Figure 4.1 (4.4.1). Distribution of staminate and pistillate flowers in the 40 first nodes of the plant. In each node, blue, red, green and yellow bars represent the percentages of male, female, bisexual and hermaphrodite flowers in the total number of plants analyzed in *etr1a* and *etr2b* mutant families (30 plants for each genotype).

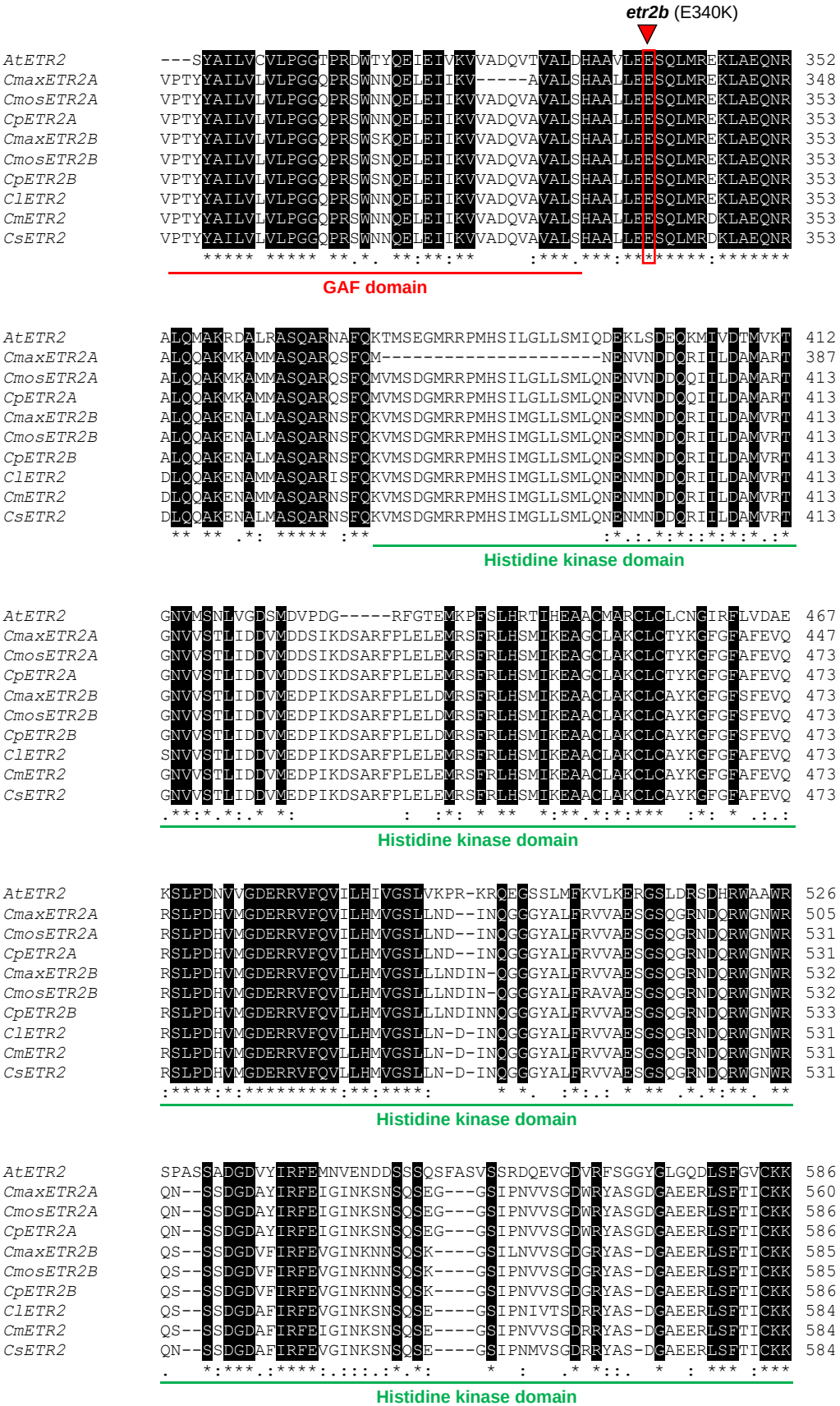

AtETR1
CmaxETR1A
CpETR1A
CmosETR1A
CmaxETR1B
CpETR1B
CmosETR1B
ClETR1
CmETR1
CsETR1
Transmembrane domain
etr1a (A95V)
Transmembrane domain
Transmembrane domain
GAF domain
GAF domain
GAF domain

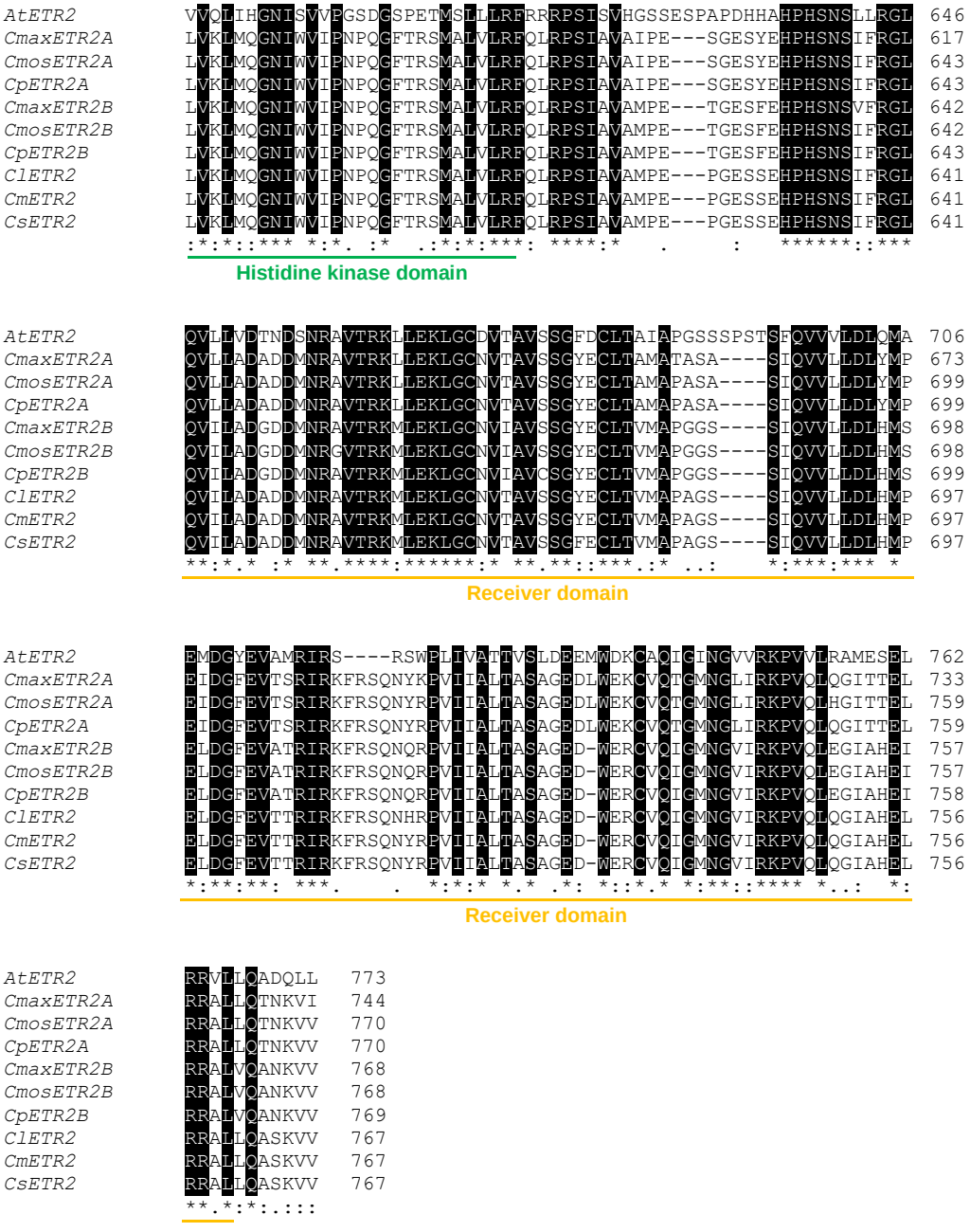
```
AtETR1      DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmaxETR1A   DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CpETR1A     DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmosETR1A   DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmaxETR1B   DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CpETR1B     DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmosETR1B   DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
ClETR1      DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmETR1      DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CsETR1      DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
            ************************************************************
GAF

AtETR1      AIIALSSLLQETELTPEQRLMVETILKSSNLLATLMNDVLDLSRLEDGSLQLELGTFNLH 420
CmaxETR1A   AVIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CpETR1A     AVIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmosETR1A   AVIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmaxETR1B   AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CpETR1B     AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmosETR1B   AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
ClETR1      AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmETR1      AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CsETR1      AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
            *:*********************************:****************::******
Histidine kinase domain

AtETR1      TLFREVLNLIKPIAVVKKLPITLNLAPDLPEFVVGDEKRLMQIILNIVGNAVKFSKQGSI 480
CmaxETR1A   AVFKEVLNLIKPVALVKKLSLTLHLGHDLPAYAIGDEKRLMQAILNVVGNAVKFSKEGSI 480
CpETR1A     AVFKEVLNLIKPVALVKKLSLTLHLGPDLPAYAIGDEKRLMQAILNVVGNAVKFSKEGSI 480
CmosETR1A   AVFKEVLNLIKPVALVKKLSLTLHLGPDLPIYAIGDEKRLMQAILNVVGNAVKFSKEGSI 480
CmaxETR1B   AVFKEVHNLIKPIALVKKLSLTLHLGPDLPVFAVGDEKRLMQTILNVVGNAVKFSKEGSI 480
CpETR1B     AVLKEVHNLIKPVALVKKLSLTLHLGPDLPVFAVGDEKRLMQTILNVVGNAVKFSKEGSI 480
CmosETR1B   AVFKEVHNLIKPVALVKKLSLTLHLGPDLPVFAVGDEKRLMQTILNVVGNAVKFSKQGSI 480
ClETR1      GVFKEVLNLIKPVALVKKLSLTLHLGPDLPVFAVGDEKRLMQAILNVVGNAVKFSKEGSI 480
CmETR1      AVFKEVLNLIKPVTLVKKLSLTLHLGPDLPVFAVGDEKRLMQAILNVVGNAVKFSKEGSI 480
CsETR1      AVFKEVLNLIKPVTLVKKLSLTLHLGLDLPVFAVGDEKRLMQAILNVVGNAVKFSKEGSI 480
             :::** ***** ::::**** :**:*. *** :.:******** ***:*********:***
Histidine kinase domain

AtETR1      SVTALVTKSDT----RAADFFVVPTGSHFYLRVKVKDSGAGINPQDIPKIFTKFAQTQSL 536
CmaxETR1A   SITAIVAKSETFREFRVQDFLPVPSDSHFYLQVQVKDAGSGISPQEIPKLFTKFAQTQTV 540
CpETR1A     SITAIVAKSETFRECRVQDFLPVPSDNHFYLRVQVKDAGSGISPQEIPKLFTKFAQTHTV 540
CmosETR1A   SITAIVAKSETFREFRAQDFLPVPSDNHFYLRVQVKDAGSGISPQEIPKLFTKFAQTQTV 540
CmaxETR1B   TITAIVAKSETFRELRLPDFHPVPSDSQFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CpETR1B     TITAIVAKSETFRELRLPDFHPAPSDSQFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CmosETR1B   TITAIVAKSETFRELRLPDFHPVPSDSQFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
ClETR1      SISAIVAKSETFREFRIPDFHPVPSDSHFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CmETR1      SISAIVAKSETFREIRVPDFHPVPSDSHFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CsETR1      SISAIVAKAETFREIRVPDFHPVPSDSHFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
            :::*:*:*::*    *  **  .*:..:*** *:***:*:**.**:***:******* ::
Histidine kinase domain

AtETR1      ATRSSGGSGLGLAISKRFVNLMEGNIWIESDGLGKGCTAIFDVKLGISERSNESKQSGIP 596
CmaxETR1A   ATRNSSCSGLGLAICKRFVNLMEGQIWLESEGLGKGCTATFIVKLGIADQSNESKLPFTS 600
CpETR1A     ATRNSSGSGLGLAICKRFVNLMEGQIWLESEGLGKGCTATFIVKLGIADQSNESKLPFTS 600
CmosETR1A   ATRNSSGSGLGLAICKRFVNLMEGQIWLESEGLGKGCTATFIVKLGIADQSNEPKLPFTS 600
CmaxETR1B   ATRNSAGSGLGLAICKRFVNLMEGHIWLESEGLGKGCLATFIVKLGIADQSNDSKLPFTS 599
CpETR1B     ATRNSAGSGLGLAICKRFVNLMEGHIWLESEGLGKGCLATFIVKLGIADQSNDSKLPFTS 599
CmosETR1B   ATRNSAGSGLGLAICKRFVNLMEGHIWLESEGLGKGCLATFIVKLGIADQSNDSKLPFTS 599
ClETR1      GPRNSDGSGLGLAICKRFVNLMEGHIWLESEGLGKGCTATFIVKLGIADQSNESKLPFTS 599
CmETR1      GPRNSGGSGLGLAICKRFVNLMEGHIWLESEGLGKGCTATFIVKLGIADQSNESKLPYTS 599
CsETR1      GPRNSCGSGLGLAICKRFVNLMEGHIWLESEGLGKGCTATFIVKLGIAEQSNESKLPFTS 599
            . *.*  *******.*********:**:**:****** * * *****:::**: *
Histidine kinase domain
```



Supplementary Figure 4.2 (4.4.5). Alignment of CpETR1A amino acid sequence with homologous sequences from diverse species. The blue lines represent the transmembrane domains, red lines represent GAF domain, green lines represent histidine kinase domains and yellow lines correspond to receiver domain. The red rectangle indicates a conserved alanine residue that has been mutated in *etr1a*.


AtETR2
CmaxETR2A
CmosETR2A
CpETR2A
CmaxETR2B
CmosETR2B
CpETR2B
ClETR2
CmETR2
CsETR2
Transmembrane domain
Transmembrane domain
Transmembrane domain
GAF domain
GAF domain


etr2b (E340K)
AtETR2 ---SYAILVCVLPGGTPRDWTYQEIEIVKVVADQVTVALDHAAVLEESQLMREKLAEQNR 352
CmaxETR2A VPTYYAILVLVLPGGQPRSWNNQELEIIKV-----AVALSHAALLEESQLMREKLAEQNR 348
CmosETR2A VPTYYAILVLVLPGGQPRSWNNQELEIIKVVADQVAVALSHAALLEESQLMREKLAEQNR 353
CpETR2A VPTYYAILVLVLPGGQPRSWNNQELEIIKVVADQVAVALSHAALLEESQLMREKLAEQNR 353
CmaxETR2B VPTYYAILVLVLPGGQPRSWSKQELEIIKVVADQVAVALSHAALLEESQLMREKLAEQNR 353
CmosETR2B VPTYYAILVLVLPGGQPRSWSNQELEIIKVVADQVAVALSHAALLEESQLMREKLAEQNR 353
CpETR2B VPTYYAILVLVLPGGQPRSWSNQELEIIKVVADQVAVALSHAALLEESQLMREKLAEQNR 353
ClETR2 VPTYYAILVLVLPGGQPRSWNNQELEIIKVVADQVAVALSHAALLEESQLMREKLAEQNR 353
CmETR2 VPTYYAILVLVLPGGQPRSWNNQELEIIKVVADQVAVALSHAALLEESQLMRDKLAEQNR 353
CsETR2 VPTYYAILVLVLPGGQPRSWNNQELEIIKVVADQVAVALSHAALLEESQLMRDKLAEQNR 353
GAF domain
AtETR2 ALQMAKRDALRASQARNAFQKTMSEGMRRPMHSILGLLSMIQDEKLSDEQKMIVDIMVKT 412
CmaxETR2A ALQQAKMKAMMASQARQSFQM--------------------NENVNDDQRILLDAMART 387
CmosETR2A ALQQAKMKAMMASQARQSFQMVMSDGMRRPMHSILGLLSMLQNENVNDDQQILLDAMART 413
CpETR2A ALQQAKMKAMMASQARQSFQMVMSDGMRRPMHSILGLLSMLQNENVNDDQQILLDAMART 413
CmaxETR2B ALQQAKENALMASQARNSFQKVMSDGMRRPMHSIMGLLSMLQNESMNDDQRILLDAMVRT 413
CmosETR2B ALQQAKENALMASQARNSFQKVMSDGMRRPMHSIMGLLSMLQNESMNDDQRILLDAMVRT 413
CpETR2B ALQQAKENALMASQARNSFQKVMSDGMRRPMHSIMGLLSMLQNESMNDDQRILLDAMVRT 413
ClETR2 DLQQAKENAMMASQARISFQKVMSDGMRRPMHSIMGLLSMLQNENMNDDQRILLDAMVRT 413
CmETR2 DLQQAKENAMMASQARNSFQKVMSDGMRRPMHSIMGLLSMLQNENMNDDQRILLDAMVRT 413
CsETR2 DLQQAKENALMASQARNSFQKVMSDGMRRPMHSIMGLLSMLQNENMNDDQRILLDAMVRT 413
Histidine kinase domain
AtETR2 GNVMSNLVGDSMDVPDG-----RFGTEMKPESLHRTIHEAAQMARCLCLCNGIRFLVDAE 467
CmaxETR2A GNVVSTLIDDVMDDSIKDSARFPLELEMRSERLHSMIKEAGQLAKCLCTYKGFGFAFEVQ 447
CmosETR2A GNVVSTLIDDVMDDSIKDSARFPLELEMRSERLHSMIKEAGQLAKCLCTYKGFGFAFEVQ 473
CpETR2A GNVVSTLIDDVMDDSIKDSARFPLELEMRSERLHSMIKEAGQLAKCLCTYKGFGFAFEVQ 473
CmaxETR2B GNVVSTLIDDVMEDPIKDSARFPLELDMRSERLHSMIKEAAQLAKCLCAYKGFGFSFEVQ 473
CmosETR2B GNVVSTLIDDVMEDPIKDSARFPLELDMRSERLHSMIKEAAQLAKCLCAYKGFGFSFEVQ 473
CpETR2B GNVVSTLIDDVMEDPIKDSARFPLELDMRSERLHSMIKEAAQLAKCLCAYKGFGFSFEVQ 473
ClETR2 SNVVSTLIDDVMEDPIKDSARFPLELEMRSERLHSMIKEAAQLAKCLCAYKGFGFAFEVQ 473
CmETR2 GNVVSTLIDDVMEDPIKDSARFPLELEMRSERLHSMIKEAAQLAKCLCAYKGFGFAFEVQ 473
CsETR2 GNVVSTLIDDVMEDPIKDSARFPLELEMRSERLHSMIKEAAQLAKCLCAYKGFGFAFEVQ 473
Histidine kinase domain
AtETR2 KSLPDNVVGDERRVFQVLLHIVGSLVKPR-KRQEGSSLMFKVLKERGSLDRSDHRWAAWR 526
CmaxETR2A RSLPDHVMGDERRVFQVLLHMVGSLLND--INQGGGYALFRVVAESGSQGRNDQRWGNWR 505
CmosETR2A RSLPDHVMGDERRVFQVLLHMVGSLLND--INQGGGYALFRVVAESGSQGRNDQRWGNWR 531
CpETR2A RSLPDHVMGDERRVFQVLLHMVGSLLND--INQGGGYALFRVVAESGSQGRNDQRWGNWR 531
CmaxETR2B RSLPDHVMGDERRVFQVLLHMVGSLLLNDIN-QGGGYALFRVVAESGSQGRNDQRWGNWR 532
CmosETR2B RSLPDHVMGDERRVFQVLLHMVGSLLLNDIN-QGGGYALFRAVAESGSQGRNDQRWGNWR 532
CpETR2B RSLPDHVMGDERRVFQVLLHMVGSLLLNDINNQGGGYALFRVVAESGSQGRNDQRWGNWR 533
ClETR2 RSLPDHVMGDERRVFQVLLHMVGSLLN-D-INQGGGYALFRVVAESGSQGRNDQRWGNWR 531
CmETR2 RSLPDHVMGDERRVFQVLLHMVGSLLN-D-INQGGGYALFRVVAESGSQGRNDQRWGNWR 531
CsETR2 RSLPDHVMGDERRVFQVLLHMVGSLLN-D-INQGGGYALFRVVAESGSQGRNDQRWGNWR 531
Histidine kinase domain
AtETR2 SPASSADGDVYIRFEMNVENDDSSSQSFASVSSRDQEVGDVRFSGGYGLGQDLSFGVCKK 586
CmaxETR2A QN--SSDGDAYIRFEIGINKSNSQSEG---GSIPNVVSGDWRYASGDGAEERLSFTICKK 560
CmosETR2A QN--SSDGDAYIRFEIGINKSNSQSEG---GSIPNVVSGDWRYASGDGAEERLSFTICKK 586
CpETR2A QN--SSDGDAYIRFEIGINKSNSQSEG---GSIPNVVSGDWRYASGDGAEERLSFTICKK 586
CmaxETR2B QS--SSDGDVFIRFEVGINKNNSQSK----GSILNVVSGDGRYAS-DGAEERLSFTICKK 585
CmosETR2B QS--SSDGDVFIRFEVGINKNNSQSK----GSIPNVVSGDGRYAS-DGAEERLSFTICKK 585
CpETR2B QS--SSDGDVFIRFEVGINKNNSQSK----GSIPNVVSGDGRYAS-DGAEERLSFTICKK 586
ClETR2 QS--SSDGDAFIRFEVGINKSNSQSE----GSIPNIVTSDRRYAS-DGAEERLSFTICKK 584
CmETR2 QS--SSDGDAFIRFEIGINKSNSQSE----GSIPNVVSGDRRYAS-DGAEERLSFTICKK 584
CsETR2 QN--SSDGDAFIRFEVGINKSNSQSE----GSIPNMVSGDRRYAS-DGAEERLSFTICKK 584
Histidine kinase domain



Supplementary Figure 4.3 (4.4.5). Alignment of CpETR2B amino acid sequence with homologous sequences from diverse species. The blue lines represent the transmembrane domains, red lines represent GAF domain, green lines represent histidine kinase domains and yellow lines correspond to receiver domain. The red rectangle indicates a conserved glutamate residue that has been mutated in *etr2b*.

Supplementary Table 5.1 (5.3.4). Primers and TaqMan probes used for genotyping *etr1b* and *etr1a-1* mutations.

etr1b family	**Sequence**	**Tm (°C)**
Forward	GGTTCTTGTTCAGTTTGGTG	56.4
Reverse	ATAATATGTACAAGCATAAGGGCA	58.3
Probe C	FAM—TTTAACCGCTGTGGTATCGTGTGCAA—BHQ1	66.2
Probe T	HEX—TTTAATCGCTGTGGTATCGTGTGCAA—BHQ1	64.2
etr1a-1 **family**		
Forward	CTCTTGAAGAAGTATCAGTACA	56.4
Reverse	ACAAAGAACAATGAAAGCACC	55.4
Probe C	FAM—TTCTCCATCCCCTTGGAGCTCA—BHQ1	64.2
Probe T	HEX—TTCTCCATCCTCTTGGAGCTCA—BHQ1	62.1

Supplementary Table 5.2 (5.3.5). Primers used for gene expression analysis by qPCR.

Gene	**Primer name**	**Sequence**
CpEf-1α	EF-1α-F	CGTCAAGAAGAAATAAGCCA
	EF-1α-R	CTACTACGAGAGAGAGAGCCG
CpETR1A	CpETR1.1-F	AAAGGAGAGCTGCCTGAGAGTC
	CpETR1.1-R	CACGACGCTCTATAAGTTCCGA
CpETR1B	CpETR1.2-F	GAGCGTCGGGTTCTATTCGA
	CpETR1.2-R	AACCTGGGATATGCCTTGTATGTAC
CpETR2B	CpETR2.1-F	TGAAGACTGGGAGAGATGC
	CpETR2.1-R1	CGGTTCCATTCATCCAAT
CpETR2A	CpETR2.2-F1	GATCAAAAAATGGCTTCTGC
	CpETR2.2-R	GAGACCTACTCGACTCGATAGA
CpACS27A	CpACS27A-F	CCAATACGGACGGTGAA
	CpACS27A-R	GGAGAAGCTGAAGAAGGAAG
CpACS11A	CpACS11A-F	CGTCGTCTTAAGGCCTTTG
	CpACS11A-R	GGTGTACCTAATTTAACGCAAC
CpACO2A	CpACO2A-F	GGAGGGAGAGGAAGATAAGG
	CpACO2A-R	CTCCATTTTCCAATAACCCA
CpWIP1B	CpWIP1.B1-F	TCCCTACTCCATGCTTCAC
	CpWIP1.B1-R	TCCTCCTCATTCACAACAAC
CpACO2B	CpACO2B-F	GGAGGGAGAGGAAGATAAGG
	CpACO2B-R	TGGGTTATTGGAAAATGGAG
CpACS27B	CpACS27B-F	AATTCATTCGGCTCAACC
	CpACS27B-R	GCTCCAGCTCCAATTTCA
CpWIP1A	CpWIP1B2-F	TGTGAAGAAGACGATGACG
	CpWIP1B2-R	ACGATGACGAGACCGAGTA

Supplementary Table 5.3 (5.3.6). Proteins used to perform the phylogenetic analysis.

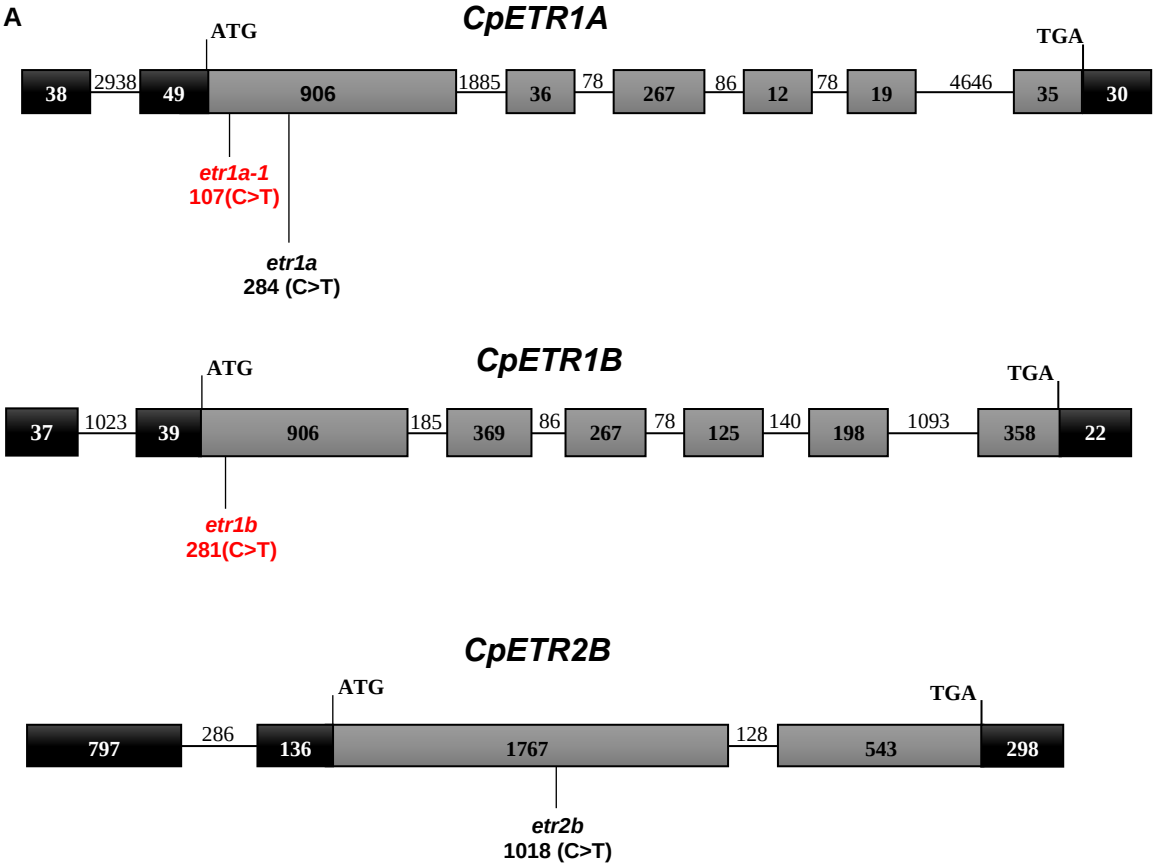
Species	Gene Name	Abbreviation
Arabidopsis thaliana	AT3G57670	***AtWIP2***
	AT4G26200	***AtACS7***
	AT4G08040	***AtACS11***
	AT1G05010	***AtACO2***
Cucumis sativus	Csa4G290830	***CsWIP1***
	Csa1G580750	***CsACS2***
	ClCG08G016770	***CsACS11***
	Csa6G511860	***CsACO2***
Cucumis melo	ACX85639	***CmWIP1***
	MELO3C015444	***CmACS7***
	MELO3C010779	***CmACS11***
	MELO3C007425	***CmACO2***
Citrullus lanatus	ClCG02G023350	***ClWIP1***
	ClCG03G015830	***ClACS4***
	Csa2G353460	***ClACS11***
	ClCG01G021590	***ClACO2***
Cucurbita pepo	Cp4.1LG05g05020	***CpWIP1A***
	Cp4.1LG16g08860	***CpWIP1B***
	Cp4.1LG18g03790	***CpACS27A***
	Cp4.1LG04g10620	***CpACS27B***
	Cp4.1LG11g01010	***CpACS11***
	Cp4.1LG19g08030	***CpACO2A***
	Cp4.1LG10g09730	***CpACO2B***
Cucurbita moschata	CmoCh02G012790	***CmosWIP1A***
	CmoCh20G000810	***CmosWIP1B***
	CmoCh11G006870	***CmosACS27A***
	CmoCh10G007280	***CmosACS27B***
	CmoCh05G001210	***CmosACS11***
	CmoCh07G003260	***CmosACO2A***
	CmoCh03G005880	***CmosACO2B***
Cucurbita maxima	CmaCh02G012400	***CmaxWIP1A***
	CmaCh20G000750	***CmaxWIP1B***
	CmaCh11G006760	***CmaxACS27A***
	CmaCh10G007020	***CmaxACS27B***
	CmaCh05G001180	***CmaxACS11***
	CmaCh07G003370	***CmaxACO2A***
	CmaCh03G005620	***CmaxACO2B***

Supplementary Table 5.4 (5.4.1). Inheritance of the *etr1a-1* and *etr1b* ethylene insensitive mutations in the backcrossing and selfing generations.

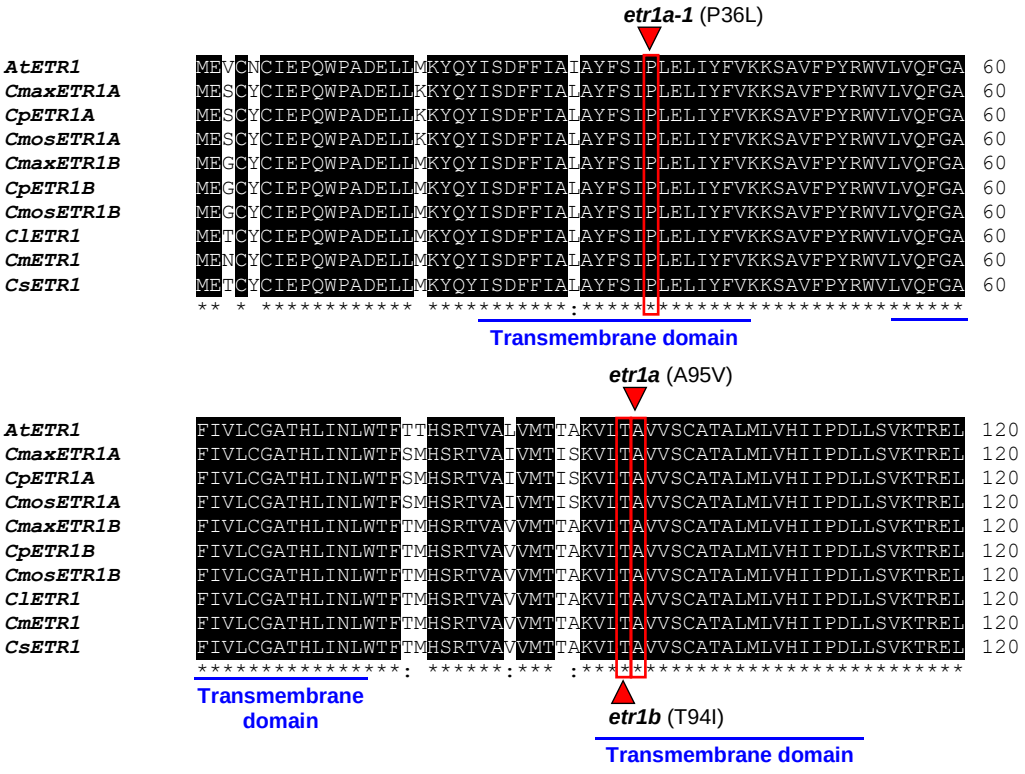
	Generation	Number of plants			Expected segregation	χ^2	p-value
		Ethylene sensitive (WT)	Intermediate (wt/*etr*)	Ethylene insensitive (*etr*/*etr*)			
etr1a-1	BC_1	40	-	42	1:0:1	0.05	0.83
	BC_2	47	-	43	1:0:1	0.18	0.67
	BC_3	125	-	119	1:0:1	0.15	0.7
	BC_4	106	-	109	1:0:1	0.04	0.84
etr1b	BC_1	-	89	-	0:1:0	-	-
	BC_2	41	39	-	1:1:0	0.05	0.82
	BC_3	56	62		1:1:0	0.31	0.58
	BC_1S_1	130	256	119	1:2:1	0.58	0.75
	BC_2S_1	96	202	89	1:2:1	1	0.61
	BC_3S_1	121	250	128	1:2:1	0.2	0.91

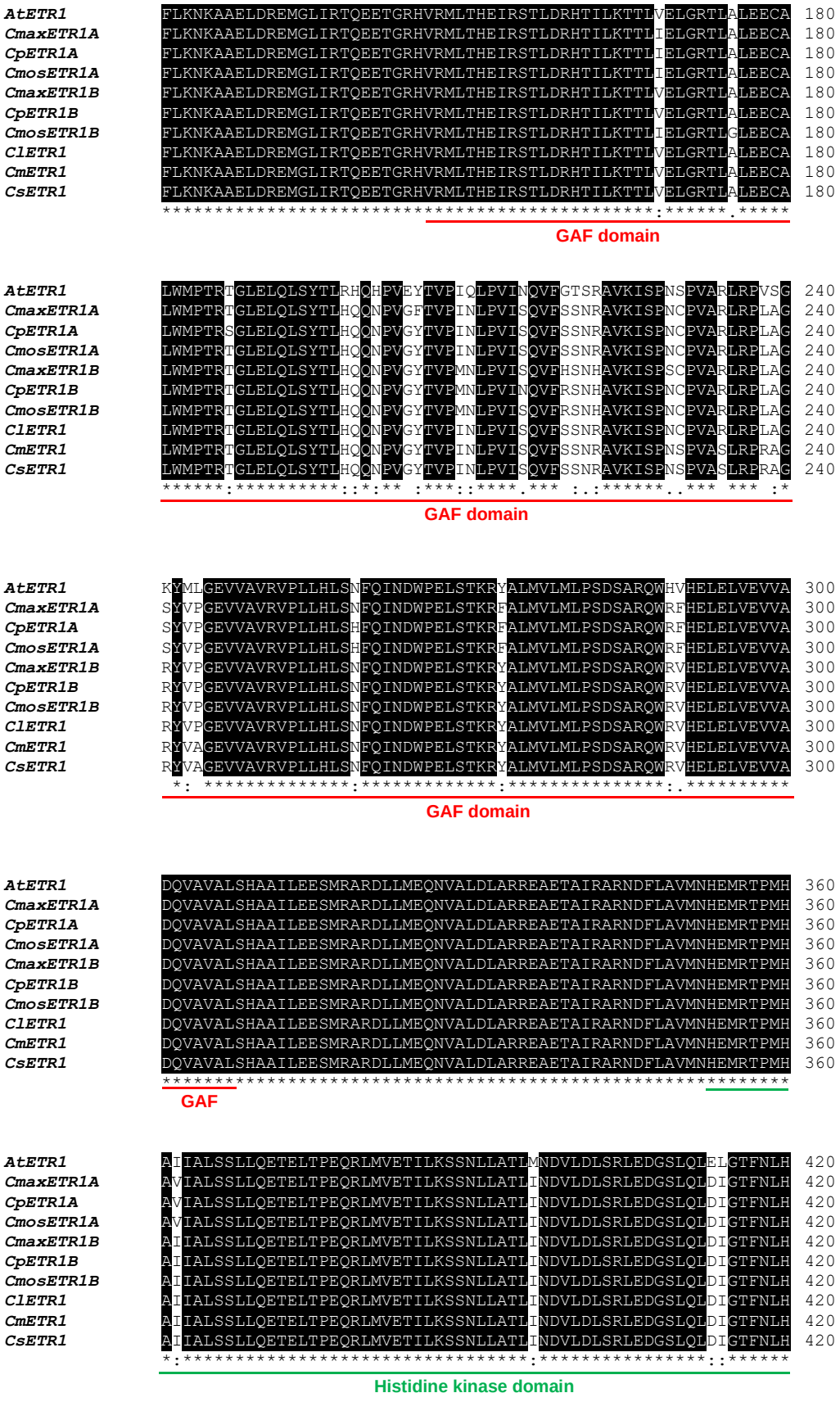
Supplementary Table 5.5 (5.4.3). Sequence and coverage statistics. Number of sequences, mapped reads and number of EMS SNPs per sample based on quality and depth, and filtering steps followed to detect causal mutations of *etr1a-1* and *etr1b* phenotypes.

	etr1b family		***etr1a-1 family***	
	WT- bulk	**Mutant-bulk**	**WT-bulk**	**Mutant-bulk**
Phenotype	Monoecious	Androecious	Monoecious	Androecious
C. pepo reference genome	C.pepo_v4.1	C.pepo_v4.1	C.pepo_v4.1	C.pepo_v4.1
No of paired reads	25,370,315	20,668,614	46,701,758	50,868,716
% mapped reads	91.45	91.27	97.91	97.61
Mean depth	13.89	12.12	23.43	25.06
EMS SNPs family specific (CT>GA)	14478	14284	11798	11598
EMS SNPs (GQ > 20; DP > 10)	521	554	4246	4056
EMS (0.3 < Ref. Freq. > 0.7)	341	340	1576	1601
SNP index (WT-bulk=0)		4		31
SNP index (*etr1b*-bulk) = 0.9 - 1		1		-
SNP index (*etr1a-1*-bulk) = 0.4 - 0.6		-		14
High impact SNPs		1		2


A
CpETR1A
ATG
TGA
38
2938
49
906
1885
36
78
267
86
12
78
19
4646
35
30
etr1a-1
107(C>T)
etr1a
284 (C>T)
CpETR1B
ATG
TGA
37
1023
39
906
185
369
86
267
78
125
140
198
1093
358
22
etr1b
281(C>T)
CpETR2B
ATG
TGA
797
286
136
1767
128
543
298
etr2b
1018 (C>T)

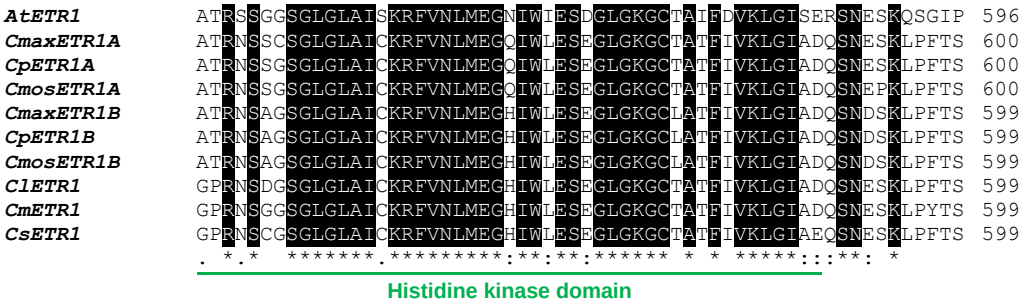
B


etr1a-1 (P36L)
AtETR1
CmaxETR1A
CpETR1A
CmosETR1A
CmaxETR1B
CpETR1B
CmosETR1B
ClETR1
CmETR1
CsETR1
Transmembrane domain
etr1a (A95V)
Transmembrane domain
etr1b (T94I)
Transmembrane domain

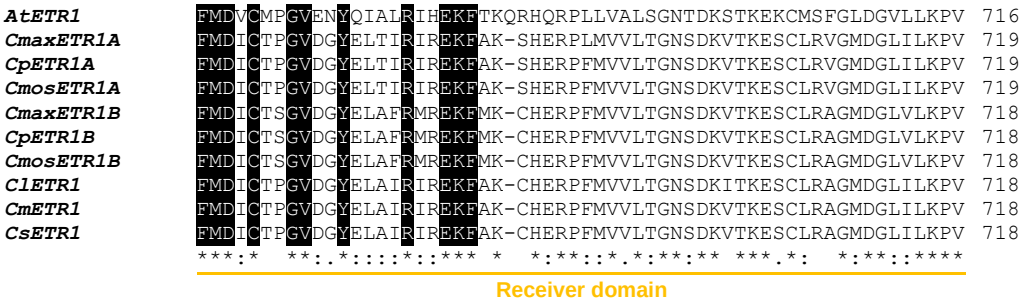

AtETR1 FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLVELGRTLALEECA 180
CmaxETR1A FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLIELGRTLALEECA 180
CpETR1A FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLIELGRTLALEECA 180
CmosETR1A FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLIELGRTLALEECA 180
CmaxETR1B FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLVELGRTLALEECA 180
CpETR1B FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLVELGRTLALEECA 180
CmosETR1B FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLIELGRTLGLEECA 180
ClETR1 FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLVELGRTLALEECA 180
CmETR1 FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLVELGRTLALEECA 180
CsETR1 FLKNKAAELDREMGLIRTQEETGRHVRMLTHEIRSTLDRHTILKTTLVELGRTLALEECA 180
GAF domain
AtETR1 LWMPTRTGLELQLSYTLRHQHPVEYTVPIQLPVINQVFGTSRAVKISPNSPVARLRPVSG 240
CmaxETR1A LWMPTRTGLELQLSYTLHQQNPVGFTVPINLPVISQVFSSNRAVKISPNCPVARLRPLAG 240
CpETR1A LWMPTRSGLELQLSYTLHQQNPVGYTVPINLPVISQVFSSNRAVKISPNCPVARLRPLAG 240
CmosETR1A LWMPTRTGLELQLSYTLHQQNPVGYTVPINLPVISQVFSSNRAVKISPNCPVARLRPLAG 240
CmaxETR1B LWMPTRTGLELQLSYTLHQQNPVGYTVPMNLPVISQVFHSNHAVKISPSCPVARLRPLAG 240
CpETR1B LWMPTRTGLELQLSYTLHQQNPVGYTVPMNLPVINQVFRSNHAVKISPNCPVARLRPLAG 240
CmosETR1B LWMPTRTGLELQLSYTLHQQNPVGYTVPMNLPVISQVFRSNHAVKISPNCPVARLRPLAG 240
ClETR1 LWMPTRTGLELQLSYTLHQQNPVGYTVPINLPVISQVFSSNRAVKISPNCPVARLRPLAG 240
CmETR1 LWMPTRTGLELQLSYTLHQQNPVGYTVPINLPVISQVFSSNRAVKISPNSPVASLRPRAG 240
CsETR1 LWMPTRTGLELQLSYTLHQQNPVGYTVPINLPVISQVFSSNRAVKISPNSPVASLRPRAG 240
GAF domain
AtETR1 KYMLGEVVAVRVPLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQWHVHELELVEVVA 300
CmaxETR1A SYVPGEVVAVRVPLLHLSNFQINDWPELSTKRFALMVLMLPSDSARQWRFHELELVEVVA 300
CpETR1A SYVPGEVVAVRVPLLHLSHFQINDWPELSTKRFALMVLMLPSDSARQWRFHELELVEVVA 300
CmosETR1A SYVPGEVVAVRVPLLHLSHFQINDWPELSTKRFALMVLMLPSDSARQWRFHELELVEVVA 300
CmaxETR1B RYVPGEVVAVRVPLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQWRVHELELVEVVA 300
CpETR1B RYVPGEVVAVRVPLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQWRVHELELVEVVA 300
CmosETR1B RYVPGEVVAVRVPLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQWRVHELELVEVVA 300
ClETR1 RYVPGEVVAVRVPLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQWRVHELELVEVVA 300
CmETR1 RYVAGEVVAVRVPLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQWRVHELELVEVVA 300
CsETR1 RYVAGEVVAVRVPLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQWRVHELELVEVVA 300
GAF domain
AtETR1 DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmaxETR1A DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CpETR1A DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmosETR1A DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmaxETR1B DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CpETR1B DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmosETR1B DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
ClETR1 DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CmETR1 DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
CsETR1 DQVAVALSHAAILEESMRARDLLMEQNVALDLARREAETAIRARNDFLAVMNHEMRTPMH 360
GAF
AtETR1 AIIALSSLLQETELTPEQRLMVETILKSSNLLATLMNDVLDLSRLEDGSLQLELGTFNLH 420
CmaxETR1A AVIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CpETR1A AVIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmosETR1A AVIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmaxETR1B AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CpETR1B AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmosETR1B AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
ClETR1 AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CmETR1 AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
CsETR1 AIIALSSLLQETELTPEQRLMVETILKSSNLLATLINDVLDLSRLEDGSLQLDIGTFNLH 420
Histidine kinase domain

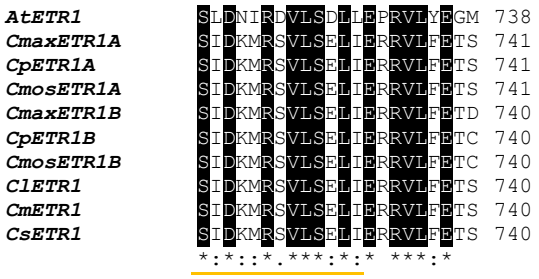

AtETR1 TLFREVLNLIKPIAVVKKLPITLNLAPDLPEFVVGDEKRLMQIILNIVGNAVKFSKQGSI 480
CmaxETR1A AVFKEVLNLIKPVALVKKLSLTLHLGHDLPAYAIGDEKRLMQAILNVVGNAVKFSKEGSI 480
CpETR1A AVFKEVLNLIKPVALVKKLSLTLHLGPDLPAYAIGDEKRLMQAILNVVGNAVKFSKEGSI 480
CmosETR1A AVFKEVLNLIKPVALVKKLSLTLHLGPDLPIYAIGDEKRLMQAILNVVGNAVKFSKEGSI 480
CmaxETR1B AVFKEVHNLIKPIALVKKLSLTLHLGPDLPVFAVGDEKRLMQTILNVVGNAVKFSKEGSI 480
CpETR1B AVLKEVHNLIKPVALVKKLSLTLHLGPDLPVFAVGDEKRLMQTILNVVGNAVKFSKEGSI 480
CmosETR1B AVFKEVHNLIKPVALVKKLSLTLHLGPDLPVFAVGDEKRLMQTILNVVGNAVKFSKQGSI 480
ClETR1 GVFKEVLNLIKPVALVKKLSLTLHLGPDLPVFAVGDEKRLMQAILNVVGNAVKFSKEGSI 480
CmETR1 AVFKEVLNLIKPVTLVKKLSLTLHLGPDLPVFAVGDEKRLMQAILNVVGNAVKFSKEGSI 480
CsETR1 AVFKEVLNLIKPVTLVKKLSLTLHLGLDLPVFAVGDEKRLMQAILNVVGNAVKFSKEGSI 480
Histidine kinase domain


AtETR1 SVTALVTKSDT----RAADFFVVPTGSHFYLRVKVKDSGAGINPQDIPKIFTKFAQTQSL 536
CmaxETR1A SITAIVAKSETFREFRVQDFLPVPSDSHFYLCVQVKDAGSGISPQEIPKLFTKFAQTQTV 540
CpETR1A SITAIVAKSETFRECRVQDFLPVPSDNHFYLRVQVKDAGSGISPQEIPKLFTKFAQTHTV 540
CmosETR1A SITAIVAKSETFREFRAQDFLPVPSDNHFYLRVQVKDAGSGISPQEIPKLFTKFAQTQTV 540
CmaxETR1B TITAIVAKSETFRELRLPDFHPVPSDSQFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CpETR1B TITAIVAKSETFRELRLPDFHPAPSDSQFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CmosETR1B TITAIVAKSETFRELRLPDFHPVPSDSQFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
ClETR1 SISAIVAKSETFREFRIPDFHPVPSDSHFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CmETR1 SISAIVAKSETFREIRVPDFHPVPSDSHFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
CsETR1 SISAIVAKAETFREIRVPDFHPVPSDSHFYLRVQVKDTGSGISPQDIPKLFTKFAQT-TV 539
Histidine kinase domain

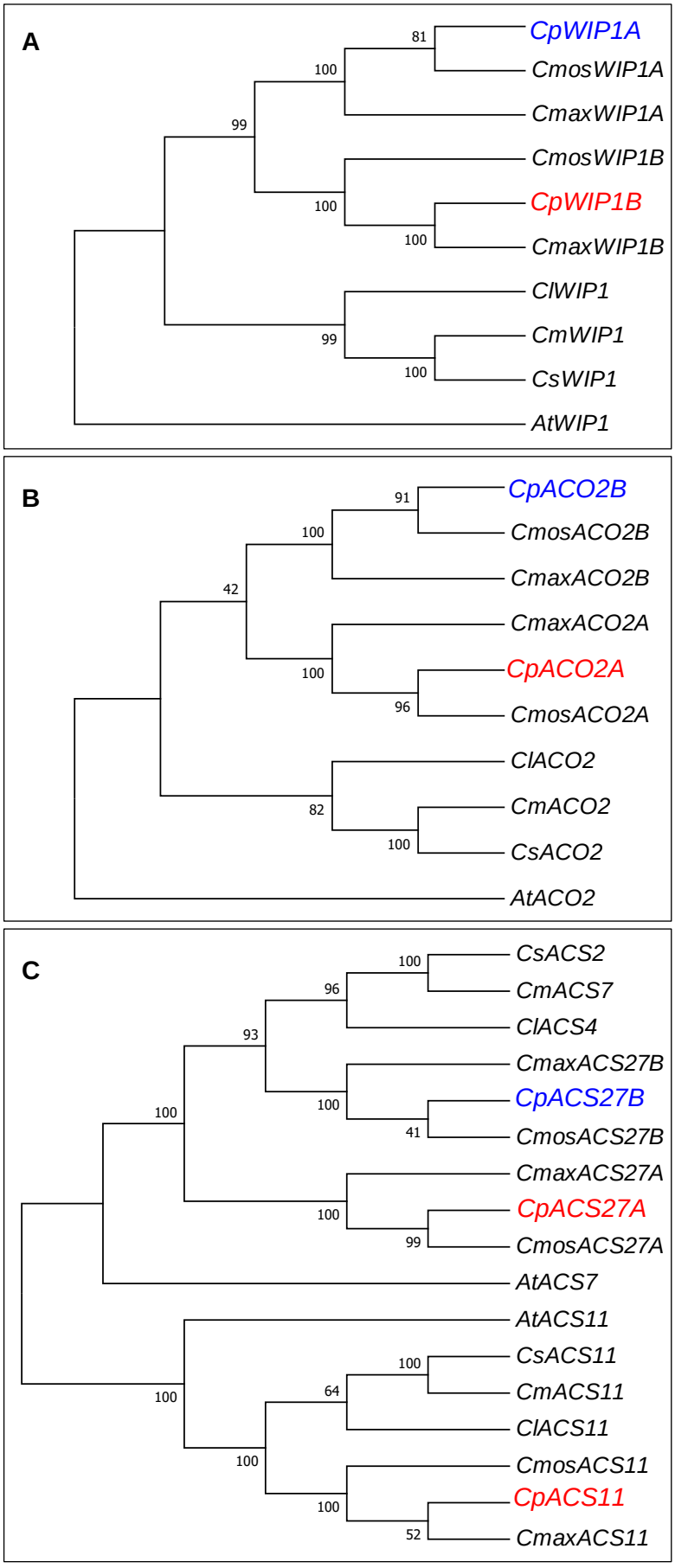

AtETR1 ATRSSGGSGLGLAISKRFVNLMEGNIWIESDGLGKGCTAIFDVKLGISERSNESKQSGIP 596
CmaxETR1A ATRNSSCSGLGLAICKRFVNLMEGQIWLESEGLGKGCTATFIVKLGIADQSNESKLPFTS 600
CpETR1A ATRNSSGSGLGLAICKRFVNLMEGQIWLESEGLGKGCTATFIVKLGIADQSNESKLPFTS 600
CmosETR1A ATRNSSGSGLGLAICKRFVNLMEGQIWLESEGLGKGCTATFIVKLGIADQSNEPKLPFTS 600
CmaxETR1B ATRNSAGSGLGLAICKRFVNLMEGHIWLESEGLGKGCLATFIVKLGIADQSNDSKLPFTS 599
CpETR1B ATRNSAGSGLGLAICKRFVNLMEGHIWLESEGLGKGCLATFIVKLGIADQSNDSKLPFTS 599
CmosETR1B ATRNSAGSGLGLAICKRFVNLMEGHIWLESEGLGKGCLATFIVKLGIADQSNDSKLPFTS 599
ClETR1 GPRNSDGSGLGLAICKRFVNLMEGHIWLESEGLGKGCTATFIVKLGIADQSNESKLPFTS 599
CmETR1 GPRNSGGSGLGLAICKRFVNLMEGHIWLESEGLGKGCTATFIVKLGIADQSNESKLPYTS 599
CsETR1 GPRNSCGSGLGLAICKRFVNLMEGHIWLESEGLGKGCTATFIVKLGIAEQSNESKLPFTS 599
Histidine kinase domain


AtETR1 KVPAIPRHSNFTGLKVLVMDENGVSRMVTKGLLVHLGCEVTTVSSNEECLRVVSHEHKVV 656
CmaxETR1A KINENNINASFPGLKVLVMDENGVSRSVTKGLLVRLGCEVTTAGSVEEFLRVVSQEHRVV 660
CpETR1A KIHENNINASFPGLKVLVMDENGVSRSVTKGLLVHLGCEVTTAGSVEEFLRVVSQEHRIV 660
CmosETR1A KIHENNINASFPGLKVLVMDENGVSRSVTKGLLVHLGCEVTTAGSVEEFLRVVSQEHRVV 660
CmaxETR1B KIHENNVHLSFPGLKVLVMEENGVSRSVTKGLLVHLGCEVTTAGSIEEFLRVVSQEHKVI 659
CpETR1B KIHENNVHPSFPGLKVLVMEENGVSRSVTKGLLVHLGCEVTTAGSIEEFLRVVSQEHKVI 659
CmosETR1B KIHENNVQPSFPGLKVLVMEENGVSRSVTKGLLVHLGCEVTTAGSIEEFLRVVSQEHKVI 659
ClETR1 KIHENNIHTSFPGLKVLVMDDNGVSRSVTKGLLVHLGCEVTTAGSIEEFLRVVSQEHKVV 659
CmETR1 KIHENSIHTSFPGLKVLVMDDNGVSRSVTKGLLVHLGCEVTTAGSIEEFLRVVSQEHKVV 659
CsETR1 KIHENSIHTSFPGLKVLVMDDNGVSRSVTKGLLVHLGCEVTTAGSIEEFLRVVSQEHKVV 659
Receiver domain


AtETR1 FMDVCMPGVENYQIALRIHEKFTKQRHQRPLLVALSGNTDKSTKEKCMSFGLDGVLLKPV 716
CmaxETR1A FMDICTPGVDGYELTIRIREKFAK-SHERPLMVVLTGNSDKVTKESCLRVGMDGLILKPV 719
CpETR1A FMDICTPGVDGYELTIRIREKFAK-SHERPFMVVLTGNSDKVTKESCLRVGMDGLILKPV 719
CmosETR1A FMDICTPGVDGYELTIRIREKFAK-SHERPFMVVLTGNSDKVTKESCLRVGMDGLILKPV 719
CmaxETR1B FMDICTSGVDGYELAFRMREKFMK-CHERPFMVVLTGNSDKVTKESCLRAGMDGLVLKPV 718
CpETR1B FMDICTSGVDGYELAFRMREKFMK-CHERPFMVVLTGNSDKVTKESCLRAGMDGLVLKPV 718
CmosETR1B FMDICTSGVDGYELAFRMREKFMK-CHERPFMVVLTGNSDKVTKESCLRAGMDGLVLKPV 718
ClETR1 FMDICTPGVDGYELAIRIREKFAK-CHERPFMVVLTGNSDKITKESCLRAGMDGLILKPV 718
CmETR1 FMDICTPGVDGYELAIRIREKFAK-CHERPFMVVLTGNSDKVTKESCLRAGMDGLILKPV 718
CsETR1 FMDICTPGVDGYELAIRIREKFAK-CHERPFMVVLTGNSDKVTKESCLRAGMDGLILKPV 718
Receiver domain



Supplementary Figure 5.1 (5.4.3). (**A**) Molecular structure of *CpETR1A*, *CpETR1B* and *CpETR2B* genes, indicating the position of 4 different ethylene insensitive mutations of *C. pepo*: *etr1a-1* and *etr1b*, characterised in this paper, and *etr1a* and *etr2b*, previously described in García et al (2019). (**B**) Alignment of CpETR1A and CpETR1B amino acid sequences with homologous ETR1 proteins from diverse species. The blue lines represent the transmembrane domains, red lines represent GAF domain, green lines represent histidine kinase domains and yellow lines correspond to receiver domain. The red rectangle indicates that *etr1a*, *etr1a-1* and *etr1b* mutations affect to very well conserved residues in ETR1 ethylene receptor.



Supplementary Figure 5.2 (5.4.5). Phylogenetic relationships between genes involves in sex determination in cucurbit species (*C. pepo*, *C. moschata*, *C. maxima*, *C. melo*, *C. sativus* and *C. lanatus*). A homologous gene from Arabidopsis was used as an outgroup. Genes with expression detected by qPCR analysis are indicated in red. Genes analyzed by qPCR with no expression detected are indicated in blue.

8. References

Abeles, F. B., Morgan, P. W., and Saltveit, M. E. (2012). *Ethylene in Plant Biology: Second Edition*. Elsevier Inc. doi:10.1016/C2009-0-03226-7.

Abeles, F. B., Morgan, P. W., and Saltveit, M. E. (1992). *Ethylene in plant biology*. Academic Press.

Adams, D. O., and Yang, S. F. (1979). Ethylene biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene. *Proc. Natl. Acad. Sci. USA.* 76, 170–4. doi:10.1073/pnas.76.1.170.

Alexander, L., and Grierson, D. (2002). Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. *J. Exp. Bot.* 53, 2039–2055. doi:10.1093/jxb/erf072.

Alonso, J. M., and Ecker, J. R. (2006). Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat. Rev. Genet.* 7, 524–36. doi:10.1038/nrg1893.

Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S., and Ecker, J. R. (1999). EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science* 284, 2148–2152. doi:10.1126/science.284.5423.2148.

Ando, S., Sato, Y., Kamachi, S., and Sakai, S. (2001). Isolation of a MADS-box gene (*ERAF17*) and correlation of its expression with the induction of formation of female flowers by ethylene in cucumber plants (*Cucumis sativus* L.). *Planta* 213, 943–52.

Argueso, C. T., Hansen, M., and Kieber, J. J. (2007). Regulation of Ethylene Biosynthesis. *J. Plant Growth Regul.* 26, 92–105. doi:10.1007/s00344-007-0013-5.

Bai, S.-L., Peng, Y.-B., Cui, J.-X., Gu, H.-T., Xu, L.-Y., Li, Y.-Q., et al. (2004). Developmental analyses reveal early arrests of the spore-bearing parts of reproductive organs in unisexual flowers of cucumber (*Cucumis sativus* L.). *Planta* 220, 230–240. doi:10.1007/s00425-004-1342-2.

Barry, C. S., Llop-Tous, M. I., and Grierson, D. (2000). The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato. *Plant Physiol.* 123, 979–86. doi:10.1104/pp.123.3.979.

Binder, B. M., and Bleecker, A. B. (2003). A model for ethylene receptor function and 1-methylcyclopropene action. *Acta Hortic.* 177–187. doi:10.17660/ActaHortic.2003.628.21.

Bisognin, D. A. (2002). Origin and evolution of cultivated cucurbits. *Ciência Rural* 32, 715–723. doi:10.1590/s0103-84782002000400028.

Bleecker, A. B., Esch, J. J., Hall, A. E., Rodriguez, F. I., and Binder, B. M. (1998). The ethylene-receptor family from Arabidopsis: structure and function. *Philos. Trans. R. Soc. B Biol. Sci.* 353, 1405–1412. doi:10.1098/rstb.1998.0295.

Bleecker, A. B., Estelle, M. A., Somerville, C., and Kende, H. (1988). Insensitivity to ethylene conferred by a dominant mutation in *Arabidopsis thaliana*. *Science* 241, 1086–1089. doi:10.1126/science.241.4869.1086.

Bleecker, A. B., and Kende, H. (2000). Ethylene: a gaseous signal molecule in plants. *Annu. Rev. Cell Dev. Biol.* 16, 1–18. doi:10.1146/annurev.cellbio.16.1.1.

Blumenstiel, J. P., Noll, A. C., Griffiths, J. A., Perera, A. G., Walton, K. N., Gilliland, W. D., et al. (2009). Identification of EMS-induced mutations in *Drosophila melanogaster* by whole-genome sequencing. *Genetics* 182, 25–32. doi:10.1534/genetics.109.101998.

Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., Morin, H., et al. (2008). A

conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* 321, 836–838. doi:10.1126/science.1159023.

Boualem, A., Fleurier, S., Troadec, C., Audigier, P., Kumar, A. P. K., Chatterjee, M., et al. (2014). Development of a *Cucumis sativus* TILLinG platform for forward and reverse genetics. *PLoS One* 9, e97963. doi:10.1371/journal.pone.0097963.

Boualem, A., Lemhemdi, A., Sari, M. A., Pignoly, S., Troadec, C., Choucha, F. A., et al. (2016). The andromonoecious sex determination gene predates the separation of *Cucumis* and *Citrullus* genera. *PLoS One* 11. doi:10.1371/journal.pone.0155444.

Boualem, A., Troadec, C., Camps, C., Lemhemdi, A., Morin, H., Sari, M.-A., et al. (2015). A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* 350, 688–691. doi:10.1126/science.aac8370.

Boualem, A., Troadec, C., Kovalski, I., Sari, M. A., Perl-Treves, R., and Bendahmane, A. (2009). A conserved ethylene biosynthesis enzyme leads to andromonoecy in two *Cucumis* species. *PLoS One* 4. doi:10.1371/journal.pone.0006144.

Byers, R. E., Baker, L. R., Dilley, D. R., and Sell, H. M. (1972a). Chemical induction of perfect flowers on a gynoecious line of muskmelon, *Cucumis melo* L. *Hortscience* 9, 321–331.

Byers, R. E., Baker, L. R., Sell, H. M., Herner, R. C., and Dilley, D. R. (1972b). Ethylene: a natural regulator of sex expression of *Cucumis melo* L. *Proc. Natl. Acad. Sci. USA.* 69, 717–720.

Caldwell, D. G., McCallum, N., Shaw, P., Muehlbauer, G. J., Marshall, D. F., and Waugh, R. (2004). A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). *Plant J.* 40, 143–150. doi:10.1111/j.1365-313X.2004.02190.x.

Cancel, J. D., and Larsen, P. B. (2002). Loss-of-function mutations in the ethylene receptor ETR1 cause enhanced sensitivity and exaggerated response to ethylene in Arabidopsis. doi:10.1104/pp.003780.

Chang, C., Kwok, S. F., Bleecker, A. B., and Meyerowitz, E. M. (1993). Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. *Science* 262, 539–44. doi:10.1126/SCIENCE.8211181.

Chang, C., and Stadler, R. (2001). Ethylene hormone receptor action in Arabidopsis. *Bioessays* 23, 619–27. doi:10.1002/bies.1087.

Chao, Q., Rothenberg, M., Solano, R., Roman, G., Terzaghi, W., and Ecker, J. R. (1997). Activation of the ethylene gas response pathway in Arabidopsis by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. *Cell* 89, 1133–44. doi:10.1016/s0092-8674(00)80300-1.

Che, G., and Zhang, X. (2019). Molecular basis of cucumber fruit domestication. *Curr. Opin. Plant Biol.* 47, 38–46. doi:10.1016/j.pbi.2018.08.006.

Chen, H., Sun, J., Li, S., Cui, Q., Zhang, H., Xin, F., et al. (2016). An ACC oxidase gene essential for cucumber carpel development. *Mol. Plant* 9, 1315–1327. doi:10.1016/j.molp.2016.06.018.

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46, 714–721. doi:10.1038/ng.3007.

Chen, Y.-F., Shakeel, S. N., Bowers, J., Zhao, X.-C., Etheridge, N., and Schaller, G. E. (2007). Ligand-induced degradation of the ethylene receptor ETR2 through a proteasome-dependent pathway in Arabidopsis. *J. Biol. Chem.* 282, 24752–8. doi:10.1074/jbc.M704419200.

Chen, Y.-H. (2012). *CsACO4*, an ACC oxidase gene regulating male differentiation in cucumber. *African J. Biotechnol.* 11. doi:10.5897/ajb09.1581.

Chen, Y. F., Gao, Z., Kerris, R. J., Wang, W., Binder, B. M., and Schaller, G. E. (2010). Ethylene receptors function as components of high-molecular-mass protein complexes in Arabidopsis. *PLoS One* 5. doi:10.1371/journal.pone.0008640.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* 6, 80–92. doi:10.4161/fly.19695.

Clark, K. L., Larsen, P. B., Wang, X., and Chang, C. (1998). Association of the Arabidopsis CTR1 Raf-like kinase with the ETR1 and ERS ethylene receptors. *Proc. Natl. Acad. Sci. USA.* 95, 5401–5406. doi:10.1073/pnas.95.9.5401.

Colbert, T., Till, B. J., Tompa, R., Reynolds, S., Steine, M. N., Yeung, A. T., et al. (2001). High-throughput screening for induced point mutations. *Plant Physiol.* 126, 480–4. doi:10.1104/PP.126.2.480.

Cooper, J. L., Till, B. J., Laport, R. G., Darlow, M. C., Kleffner, J. M., Jamai, A., et al. (2008). TILLING to detect induced mutations in soybean. *BMC Plant Biol.* 8, 9. doi:10.1186/1471-2229-8-9.

cucurbitgenomics.org Available at: http://cucurbitgenomics.org/.

Dahmani-Mardas, F., Troadec, C., Boualem, A., Lé Vê Que, S., Alsadon, A. A., Aldoss, A. A., et al. (2010). Engineering melon plants with improved fruit shelf life using the TILLING approach. *PLoS One* 5, e15776. doi:10.1371/journal.pone.0015776.

Dalmais, M., Schmidt, J., Le Signor, C., Moussy, F., Burstin, J., Savois, V., et al. (2008). UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol.* 9, R43. doi:10.1186/gb-2008-9-2-r43.

Dellaporta, S. L., and Calderon-Urrea, A. (1993). Sex determination in flowering plants. *Plant Cell* 5, 1241–51. doi:10.1105/tpc.5.10.1241.

Den Nijs, A. P. M., and Visser, D. L. (1980). Induction of male flowering in gynoecious cucumbers (*Cucumis sativus* L.) by silver ions. *Euphytica* 29, 273–280. doi:10.1007/BF00025124.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi:10.1038/ng.806.

Diggle, P. K., Di Stilio, V. S., Gschwend, A. R., Golenberg, E. M., Moore, R. C., Russell, J. R. W., et al. (2011). Multiple developmental processes underlie sex differentiation in angiosperms. *Trends Genet.* 27, 368–76. doi:10.1016/j.tig.2011.05.003.

Drake, J. W., and Baltz, R. H. (1976). The Biochemistry of Mutagenesis. *Annu. Rev. Biochem.* 45, 11–37. doi:10.1146/annurev.bi.45.070176.000303.

Esteras, C., Gomez, P., Monforte, A. J., Blanca, J., Vicente-Dolera, N., Roig, C., et al. (2012). High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. *BMC Genomics* 13, 80. doi:10.1186/1471-2164-13-80.

FAOSTAT Available at: http://www.fao.org/faostat/en/#home.

Ferriol, M., Picó, B., and Nuez, F. (2003). Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. *Theor. Appl. Genet.* 107, 271–282. doi:10.1007/s00122-

003-1242-z.

Fraenkel, R., Kovalski, I., Troadec, C., Bendahmane, A., and Perl-Treves, R. (2014). Development and evaluation of a cucumber TILLING population. *BMC Res. Notes* 7, 846. doi:10.1186/1756-0500-7-846.

Galun, E. (1959). The cucumber tendril - a new test organ for gibberellic acid. *Experientia* 15, 184–185. doi:10.1007/BF02158690.

Galun, E. (1962). Study of the inheritance of sex expression in the cucumber. The interaction of major genes with modifying genetic and non-genetic factors. *Genetica* 32, 134–163. doi:10.1007/BF01816091.

Galun, E., Izhar, S., and Atsmon, D. (1965). Determination of relative auxin content in hermaphrodite and andromonoecious *Cucumis sativus* L. *Plant Physiol.* 40, 321–326. doi:10.1104/pp.40.2.321.

Gamble, R. L., Coonfield, M. L., and Schaller, G. E. (1998). Histidine kinase activity of the ETR1 ethylene receptor from Arabidopsis. *Proc. Natl. Acad. Sci.* 95.

Gamble, R. L., Qu, X., and Schaller, G. E. (2002). Mutational analysis of the ethylene receptor ETR1. Role of the histidine kinase domain in dominant ethylene insensitivity. *PLANT Physiol.* 128, 1428–1438. doi:10.1104/pp.010777.

Gao, Z., Chen, Y.-F., Randlett, M. D., Zhao, X.-C., Findell, J. L., Kieber, J. J., et al. (2003). Localization of the Raf-like kinase CTR1 to the endoplasmic reticulum of Arabidopsis through participation in ethylene receptor signaling complexes. *J. Biol. Chem.* 278, 34725–32. doi:10.1074/jbc.M305548200.

Gao, Z., and Schaller, G. E. (2009). The role of receptor interactions in regulating ethylene signal transduction. *Plant Signal. Behav.* 4, 1152–3.

Gao, Z., Wen, C.-K., Binder, B. M., Chen, Y.-F., Chang, J., Chiang, Y.-H., et al. (2008). Heteromeric interactions among ethylene receptors mediate signaling in Arabidopsis. *J. Biol. Chem.* 283, 23801–10. doi:10.1074/jbc.M800641200.

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., Gonźalez, V. M., et al. (2012). The genome of melon (Cucumis melo L.). *Proc. Natl. Acad. Sci. USA.* 109, 11872–11877. doi:10.1073/pnas.1205415109.

García, A., Aguado, E., Martínez, C., Loska, D., Beltrán, S., Valenzuela, J. L., et al. (2019). The ethylene receptors *CpETR1A* and *CpETR2B* cooperate in the control of sex determination in *Cucurbita pepo*. *J. Exp. Bot.* doi:10.1093/jxb/erz417.

García, A., Aguado, E., Parra, G., Manzano, S., Martínez, C., Megías, Z., et al. (2018). Phenomic and genomic characterization of a mutant platform in *Cucurbita pepo*. *Front. Plant Sci.* 9, 1049. doi:10.3389/fpls.2018.01049.

González, M., Xu, M., Esteras, C., Roig, C., Monforte, A. J., Troadec, C., et al. (2011). Towards a TILLING platform for functional genomics in Piel de Sapo melons. *BMC Res. Notes* 4, 289. doi:10.1186/1756-0500-4-289.

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics* 164, 731–740.

Grefen, C., Städele, K., Růzicka, K., Obrdlik, P., Harter, K., and Horák, J. (2008). Subcellular localization and in vivo interactions of the *Arabidopsis thaliana* ethylene receptor family

members. *Mol. Plant* 1, 308–320. doi:10.1093/mp/ssm015.

Guo, H., and Ecker, J. R. (2004). The ethylene signaling pathway: new insights. *Curr. Opin. Plant Biol.* 7, 40–49. doi:10.1016/J.PBI.2003.11.011.

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–8. doi:10.1038/ng.2470.

Guzmán, P., and Ecker, J. R. (1990). Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. *Plant Cell* 2, 513–23. doi:10.1105/tpc.2.6.513.

Hall, A. E., Chen, Q. G., Findell, J. L., Schaller, G. E., and Bleecker, A. B. (1999). The relationship between ethylene binding and dominant insensitivity conferred by mutant forms of the ETR1 ethylene receptor. *Plant Physiol.* 121, 291–300. doi:10.1104/PP.121.1.291.

Hall, A. E., Findell, J. L., Schaller, G. E., Sisler, E. C., and Bleecker, A. B. (2000). Ethylene perception by the ERS1 protein in Arabidopsis. *Plant Physiol.* 123, 1449–1458. doi:10.1104/pp.123.4.1449.

Harkey, A. F., Watkins, J. M., Olex, A. L., DiNapoli, K. T., Lewis, D. R., Fetrow, J. S., et al. (2018). Identification of transcriptional and receptor networks that control root responses to ethylene. *Plant Physiol.* 176, 2095–2118. doi:10.1104/pp.17.00907.

Hirayama, T., Kieber, J. J., Hirayama, N., Kogan, M., Guzman, P., Nourizadeh, S., et al. (1999). Wilson disease-related copper transporter, is required for ethylene signaling in Arabidopsis. *Cell* 97, 383–93.

Hu, B., Li, D., Liu, X., Qi, J., Gao, D., Zhao, S., et al. (2017). Engineering non-transgenic gynoecious cucumber using an improved transformation protocol and optimized CRISPR/Cas9 system. *Mol. Plant* 10, 1575–1578. doi:10.1016/J.MOLP.2017.09.005.

Hua, J., Chang, C., Sun, Q., and Meyerowitz, E. M. (1995). Ethylene insensitivity conferred by Arabidopsis ERS gene. *Science* 269, 5231.

Hua, J., and Meyerowitz, E. M. (1998). Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*. *Cell* 94, 261–271. doi:10.1016/S0092-8674(00)81425-7.

Hua, J., Sakai, H., Nourizadeh, S., Chen, Q. G., Bleecker, A. B., Ecker, J. R., et al. (1998). *EIN4* and *ERS2* are members of the putative ethylene receptor gene family in Arabidopsis. *Plant Cell* 10, 1321–32. doi:10.1105/TPC.10.8.1321.

Huang, P. L., Parks, J. E., Rottmann, W. H., and Theologis, A. (1991). Two genes encoding 1-aminocyclopropane-1-carboxylate synthase in zucchini (*Cucurbita pepo*) are clustered and similar but differentially regulated. *Proc. Natl. Acad. Sci. USA.* 88, 7021–5. doi:10.1073/pnas.88.16.7021.

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., et al. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 41, 1275–1281. doi:10.1038/ng.475.

Huang, Y., Li, H., Hutchison, C. E., Laskey, J., and Kieber, J. J. (2003). Biochemical and functional analysis of CTR1, a protein kinase that negatively regulates ethylene signaling in Arabidopsis. *Plant J.* 33, 221–233. doi:10.1046/j.1365-313X.2003.01620.x.

Jamilena, M., Mariotti, B., and Manzano, S. (2008). Plant sex chromosomes: molecular structure and function. *Cytogenet. Genome Res.* 120, 255–264. doi:10.1159/000121075.

Ji, G., Zhang, J., Zhang, H., Sun, H., Gong, G., Shi, J., et al. (2016). Mutation in the gene encoding

1-aminocyclopropane-1-carboxylate synthase 4 (*CitACS4*) led to andromonoecy in watermelon. *J. Integr. Plant Biol.* 58, 762–765. doi:10.1111/jipb.12466.

Johnson, P. R., and Ecker, J. R. (1998). The ethylene gas signal transduction pathway: a molecular perspective. *Annu. Rev. Genet.* 32, 227–254. doi:10.1146/annurev.genet.32.1.227.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi:10.1093/bioinformatics/8.3.275.

Kahana, A., Silberstein, L., Kessler, N., Goldstein, R. S., and Perl-Treves, R. (1999). Expression of ACC oxidase genes differs among sex genotypes and sex phases in cucumber. *Plant Mol. Biol.* 41, 517–28. doi:10.1023/a:1006343707567.

Kamachi, S., Sekimoto, H., Kondo, N., and Sakai, S. (1997). Cloning of a cDNA for a 1-aminocyclopropane-1-carboxylate synthase that is expressed during development of female flowers at the apices of *Cucumis sativus* L. *Plant Cell Physiol.* 38, 1197–206. doi:10.1093/oxfordjournals.pcp.a029106.

Kastenmayer, J. P., and Green, P. J. (2000). Novel features of the XRN-family in Arabidopsis: Evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. *Proc. Natl. Acad. Sci. USA.* 97, 13985–13990. doi:10.1073/pnas.97.25.13985.

Kater, M. M., Franken, J., Carney, K. J., Colombo, L., and Angenent, G. C. (2001). Sex determination in the monoecious species cucumber is confined to specific floral whorls. *Plant Cell* 13, 481–493. doi:10.1105/tpc.13.3.481.

Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., and Ecker, J. R. (1993). CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the Raf family of protein kinases. *Cell* 72, 427–441. doi:10.1016/0092-8674(93)90119-B.

Klee, H. J. (2004). Ethylene signal transduction. Moving beyond Arabidopsis. *Plant Physiol.* 135, 660–667. doi:10.1104/pp.104.040998.

Knopf, R. R., and Trebitsh, T. (2006). The female-specific *Cs-ACS1G* gene of cucumber. A case of gene duplication and recombination between the non-sex-specific 1-aminocyclopropane-1-carboxylate synthase gene and a branched-chain amino acid transaminase gene transaminase gene. *Plant Cell Physiol.* 47, 1217–28. doi:10.1093/pcp/pcj092.

Konishi, M., and Yanagisawa, S. (2008). Two different mechanisms control ethylene sensitivity in Arabidopsis via the regulation of *EBF2* expression. *Plant Signal. Behav.* 3, 749–51. doi:10.4161/psb.3.9.6640.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054.

Kyriakis, J. M., App, H., Zhang, X. F., Banerjee, P., Brautigan, D. L., Rapp, U. R., et al. (1992). Raf-1 activates MAP kinase-kinase. *Nature* 358, 417–421. doi:10.1038/358417a0.

Lasserre, E., Bouquin, T., Hernandez, J. A., Pech, J.-C., Balagué, C., and Bull, J. (1996). Structure and expression of three genes encoding ACC oxidase homologs from melon (*Cucumis melo* L.). *Mol. Gen. Genet.* 251, 81–90. doi:10.1007/BF02174348.

Leclercq, J., Adams-Phillips, L. C., Zegzouti, H., Jones, B., Latché, A., Giovannoni, J. J., et al. (2002). *LeCTR1*, a tomato CTR1-like gene, demonstrates ethylene signaling ability in Arabidopsis and novel expression patterns in tomato. *Plant Physiol.* 130, 1132–1142. doi:10.1104/pp.009415.

Lee, J.-H., Deng, X. W., and Kim, W. T. (2006). Possible role of light in the maintenance of EIN3/EIL1 stability in Arabidopsis seedlings. *Biochem. Biophys. Res. Commun.* 350, 484–91. doi:10.1016/j.bbrc.2006.09.074.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009a). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352.

Li, Z., Huang, S., Liu, S., Pan, J., Zhang, Z., Tao, Q., et al. (2009b). Molecular isolation of the *M* gene suggests that a conserved-residue conversion induces the formation of bisexual flowers in cucumber plants. *Genetics* 182, 1381–1385. doi:10.1534/genetics.109.104737.

Li, Z., Wang, S., Tao, Q., Pan, J., Si, L., Gong, Z., et al. (2012). A putative positive feedback regulation mechanism in *CsACS2* expression suggests a modified model for sex determination in cucumber (*Cucumis sativus* L.). *J. Exp. Bot.* 63, 4475–84. doi:10.1093/jxb/ers123.

Lin, Q., Li, J., Smith, R. D., and Walker, J. C. (1998). Molecular cloning and chromosomal mapping of type one serine/threonine protein phosphatases in *Arabidopsis thaliana*. *Plant Mol. Biol.* 37, 471–481. doi:10.1023/A:1005912413555.

Lin, Z., Alexander, L., Hackett, R., and Grierson, D. (2008). LeCTR2, a CTR1-like protein kinase from tomato, plays a role in ethylene signalling, development and defence. *Plant J.* 54, 1083–1093. doi:10.1111/j.1365-313X.2008.03481.x.

Little, H. A., Papadopoulou, E., Hammar, S. A., and Grumet, R. (2007). The influence of ethylene perception on sex expression in melon (*Cucumis melo* L.) as assessed by expression of the mutant ethylene receptor, *At-etr1-1*, under the control of constitutive and floral targeted promoters. *Sex. Plant Reprod.* 20, 123–136. doi:10.1007/s00497-007-0049-5.

Lust, T. A., and Paris, H. S. (2016). Italian horticultural and culinary records of summer squash (*Cucurbita pepo, Cucurbitaceae*) and emergence of the zucchini in 19th-century Milan. *Ann. Bot.* 118, 53–69. doi:10.1093/aob/mcw080.

Malepszy, S., and Niemirowicz-Szczytt, K. (1991). Sex determination in cucumber (*Cucumis sativus*) as a model system for molecular biology. *Plant Sci.* 80, 39–47. doi:10.1016/0168-9452(91)90271-9.

Mansfeld, B. N., and Grumet, R. (2018). QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. *bioRxiv*, 208140. doi:10.1101/208140.

Manzano, S., Aguado, E., Martínez, C., Megías, Z., García, A., and Jamilena, M. (2016). The ethylene biosynthesis gene *CitACS4* regulates monoecy/andromonoecy in watermelon (*Citrullus lanatus*). *PLoS One* 11, e0154362. doi:10.1371/journal.pone.0154362.

Manzano, S., Martínez, C., Domínguez, V., Avalos, E., Garrido, D., Gómez, P., et al. (2010a). A major gene conferring reduced ethylene sensitivity and maleness in *Cucurbita pepo*. *J. Plant Growth Regul.* 29, 73–80. doi:10.1007/s00344-009-9116-5.

Manzano, S., Martínez, C., García, J. M., Megías, Z., and Jamilena, M. (2014). Involvement of ethylene in sex expression and female flower development in watermelon (*Citrullus lanatus*). *Plant Physiol. Biochem.* 85, 96–104. doi:10.1016/j.plaphy.2014.11.004.

Manzano, S., Martínez, C., Gómez, P., Garrido, D., and Jamilena, M. (2010b). Cloning and characterisation of two CTR1-like genes in Cucurbita pepo: Regulation of their expression during male and female flower development. *Sex. Plant Reprod.* 23, 301–313.

doi:10.1007/s00497-010-0140-1.

Manzano, S., Martínez, C., Megías, Z., Garrido, D., and Jamilena, M. (2013). Involvement of ethylene biosynthesis and signalling in the transition from male to female flowering in the monoecious *Cucurbita pepo*. *J. Plant Growth Regul.* 32, 789–798. doi:10.1007/s00344-013-9344-6.

Manzano, S., Martínez, C., Megías, Z., Gómez, P., Garrido, D., and Jamilena, M. (2011). The role of ethylene and brassinosteroids in the control of sex expression and flower development in *Cucurbita pepo*. *Plant Growth Regul.* 65, 213–221. doi:10.1007/s10725-011-9589-7.

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* 9, 1185–1188. doi:10.1038/nmeth.2221.

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461, 1135–1138. doi:10.1038/nature08498.

Martínez, C., Manzano, S., Megías, Z., Barrera, A., Boualem, A., Garrido, D., et al. (2014). Molecular and functional characterization of *CpACS27A* gene reveals its involvement in monoecy instability and other associated traits in squash (*Cucurbita pepo* L.). *Planta* 239, 1201–1215. doi:10.1007/s00425-014-2043-0.

Martínez, C., Manzano, S., Megías, Z., Garrido, D., Picó, B., and Jamilena, M. (2013). Involvement of ethylene biosynthesis and signalling in fruit set and early fruit development in zucchini squash (*Cucurbita pepo* L.). *BMC Plant Biol.* 13, 139. doi:10.1186/1471-2229-13-139.

Mascher, M., Jost, M., Kuon, J. E., Himmelbach, A., Abfalg, A., Beier, S., et al. (2014). Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol.* 15. doi:10.1186/gb-2014-15-6-r78.

McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123, 439–442. doi:10.1104/pp.123.2.439.

McGrath, R. B., and Ecker, J. R. (1998). Ethylene signaling in Arabidopsis: Events from the membrane to the nucleus. *Plant Physiol. Biochem.* 36, 103–113. doi:10.1016/S0981-9428(98)80095-8.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–303. doi:10.1101/gr.107524.110.

Mcmurray, A. L., and Miller, C. H. (1968). Cucumber sex expression modified by 2-chloroethanephosphonic acid. *Science* 162, 1397–1398. doi:10.1126/science.162.3860.1397.

Megías, Z., Martínez, C., Manzano, S., García, A., del Mar Rebolloso-Fuentes, M., Valenzuela, J. L., et al. (2016). Ethylene biosynthesis and signaling elements involved in chilling injury and other postharvest quality traits in the non-climacteric fruit of zucchini (*Cucurbita pepo*). *Postharvest Biol. Technol.* 113, 48–57. doi:http://dx.doi.org/10.1016/j.postharvbio.2015.11.001.

Mibus, H., and Tatlioglu, T. (2004). Molecular characterization and isolation of the *F/f* gene for femaleness in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 109, 1669–76. doi:10.1007/s00122-004-1793-7.

Minoia, S., Petrozza, A., D'Onofrio, O., Piron, F., Mosca, G., Sozio, G., et al. (2010). A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC Res. Notes* 3,

69. doi:10.1186/1756-0500-3-69.

Mirica, L. M., and Klinman, J. P. (2008). The nature of O2 activation by the ethylene-forming enzyme 1-aminocyclopropane-1-carboxylic acid oxidase. *Proc. Natl. Acad. Sci. USA.* 105, 1814–1819. doi:10.1073/pnas.0711626105.

Montero-Pau, J., Blanca, J., Bombarely, A., Ziarsolo, P., Esteras, C., Martí-Gómez, C., et al. (2017). *De novo* assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus. *Plant Biotechnol. J.* doi:10.1111/pbi.12860.

Moussatche, P., and Klee, H. J. (2004). Autophosphorylation activity of the Arabidopsis ethylene receptor multigene family. *J. Biol. Chem.* 279, 48734–41. doi:10.1074/jbc.M403100200.

Nakajima, N., Mori, H., Yamazaki, K., and Imaseki, H. (1990). Molecular cloning and sequence of a complementary DNA encoding 1-aminocyclopropane-l-carboxylate synthase induced by tissue wounding. *Plant Cell Physiol.* 31, 1021–1029. doi:10.1093/oxfordjournals.pcp.a077998.

Olmedo, G., Guo, H., Gregory, B. D., Nourizadeh, S. D., Aguilar-Henonin, L., Li, H., et al. (2006). *ETHYLENE-INSENSITIVE5* encodes a 5'-->3' exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2. *Proc. Natl. Acad. Sci. USA* 103, 13286–93. doi:10.1073/pnas.0605528103.

Ouaked, F., Rozhon, W., Lecourieux, D., and Hirt, H. (2003). A MAPK pathway mediates ethylene signaling in plants. *EMBO J.* 22, 1282–1288. doi:10.1093/emboj/cdg131.

Owens, K., KW, O., CE, P., and GE, T. (1980). Production of hermaphrodite flowers on gynoecious muskmelon by silver nitrate and aminoethoxyvinyglycine. *HortScience* 15:654-655

Pan, J., Wang, G., Wen, H., Du, H., Lian, H., He, H., et al. (2018). Differential gene expression caused by the f and m loci provides insight into ethylene-mediated female flower differentiation in cucumber. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.01091.

Pannell, J. R. (2017). Plant sex determination. *Curr. Biol.* 27, R191–R197. doi:10.1016/J.CUB.2017.01.052.

Papadopoulou, E., and Grumet, R. (2005). Brassinosteriod-induced femaleness in cucumber and relationship to ethylene production. *HortScience* 40, 1763–1767. doi:10.21273/hortsci.40.6.1763.

Paris, H. S. (1986). A proposed subspecific classification for *Cucurbita pepo*. *Phytologia* 61, 133–138.

Paris, H. S. (2000). "History of the Cultivar-Groups of *Cucurbita pepo*". *Hortic. Rev.* 71–170. doi:10.1002/9780470650783.ch2.

Paris, H. S. (2016). Germplasm enhancement of *Cucurbita pepo* (pumpkin, squash, gourd: *Cucurbitaceae*): progress and challenges. *Euphytica* 208, 415–438. doi:10.1007/s10681-015-1605-y.

Paris, H. S., Doron-Faigenboim, A., Reddy, U. K., Donahoo, R., and Levi, A. (2015). Genetic relationships in *Cucurbita pepo* (pumpkin, squash, gourd) as viewed with high frequency oligonucleotide–targeting active gene (HFO–TAG) markers. *Genet. Resour. Crop Evol.* 62, 1095–1111. doi:10.1007/s10722-015-0218-6.

Pelech, S. L., and Sanghera, J. S. (1992). MAP kinases: charting the regulatory pathways. *Science* 257, 1355–6. doi:10.1126/science.1382311.

Peñaranda, A., Payan, M. C., Garrido, D., Gómez, P., and Jamilena, M. (2007). Production of fruits

with attached flowers in zucchini squash is correlated with the arrest of maturation of female flowers. *J. Hortic. Sci. Biotechnol.* 82, 579–584. doi:10.1080/14620316.2007.11512276.

Perl-Treves, R. (2004). "Male to female conversion along the cucumber shoot: approaches to studying sex genes and floral development in *Cucumis sativus*," in *Sex Determination in Plants*, 193–221. doi:10.4324/9780203345993-13.

Peterson, C. E., and Anhder, L. D. (1960). Induction of staminate flowers on gynoecious cucumbers with gibberellin A3. *Science* 131, 1673–4. doi:10.1126/science.131.3414.1673.

Picard Tools - By Broad Institute Available at: http://broadinstitute.github.io/picard/.

Porch, T. G., Blair, M. W., Lariguet, P., Galeano, C., Pankhurst, C. E., and Broughton, W. J. (2009). Generation of a mutant population for TILLING common bean genotype BAT 93. *J. Am. Soc. Hortic. Sci.* 134, 348–355.

Potuschak, T., Lechner, E., Parmentier, Y., Yanagisawa, S., Grava, S., Koncz, C., et al. (2003). EIN3-dependent regulation of plant ethylene hormone signaling by two Arabidopsis F box proteins: EBF1 and EBF2. *Cell* 115, 679–89. doi:10.1016/s0092-8674(03)00968-1.

Potuschak, T., Vansiri, A., Binder, B. M., Lechner, E., Vierstra, R. D., and Genschik, P. (2006). The exoribonuclease XRN4 is a component of the ethylene response pathway in Arabidopsis. *Plant Cell* 18, 3047–57. doi:10.1105/tpc.106.046508.

Resnick, J. S., Rivarola, M., and Chang, C. (2008). Involvement of *RTE1* in conformational changes promoting ETR1 ethylene receptor signaling in Arabidopsis. *Plant J.* 56, 423–431. doi:10.1111/j.1365-313X.2008.03615.x.

Rodriguez, F. I., Esch, J. J., Hall, A. E., Binder, B. M., Schaller, G. E., and Bleecker, A. B. (1999). A copper cofactor for the ethylene receptor ETR1 from Arabidopsis. *Science* 283, 996–998. doi:10.1126/science.283.5404.996.

Roman, G., Lubarsky, B., Kieber, J. J., Rothenberg, M., and Ecker, J. R. (1995). Genetic analysis of ethylene signal transduction in *Arabidopsis thaliana*: five novel mutant loci integrated into a stress response pathway. *Genetics* 139, 1393–409.

Rottmann, W. H., Peter, G. F., Oeller, P. W., Keller, J. A., Shen, N. F., Nagy, B. P., et al. (1991). 1-aminocyclopropane-1-carboxylate synthase in tomato is encoded by a multigene family whose transcription is induced during fruit and floral senescence. *J. Mol. Biol.* 222, 937–61. doi:10.1016/0022-2836(91)90587-v.

Rudich, J., Halevy, A. H., and Kedar, N. (1969). Increase in femaleness of three cucurbits by treatment with Ethrel, an ethylene releasing compound. *Planta* 86, 69–76. doi:10.1007/BF00385305.

Rudich, J., Halevy, A. H., and Kedar, N. (1972a). Ethylene evolution from cucumber plants as related to sex expression. *Plant Physiol.* 49, 998–9. doi:10.1104/pp.49.6.998.

Rudich, J., Halevy, A. H., and Kedar, N. (1972b). The level of phytohormones in monoecious and gynoecious cucumbers as affected by photoperiod and ethephon. *Plant Physiol.* 50, 585–590. doi:10.1104/pp.50.5.585.

Ruduś, I., Sasiak, M., and Kepczyński, J. (2013). Regulation of ethylene biosynthesis at the level of 1-aminocyclopropane-1-carboxylate oxidase (*ACO*) gene. *Acta Physiol. Plant.* 35. doi:10.1007/s11738-012-1096-6.

Saito, S., Fujii, N., Miyazawa, Y., Yamasaki, S., Matsuura, S., Mizusawa, H., et al. (2007). Correlation between development of female flower buds and expression of the *CS-ACS2* gene

in cucumber plants. *J. Exp. Bot.* 58, 2897–2907. doi:10.1093/jxb/erm141.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-2. doi:10.1093/oxfordjournals.molbev.a040454.

Sakai, H., Hua, J., Chen, Q. G., Chang, C., Medrano, L. J., Bleecker, A. B., et al. (1998). *ETR2* is an *ETR1*-like gene involved in ethylene signaling in Arabidopsis. *Proc. Natl. Acad. Sci.* 95, 5812–5817. doi:10.1073/pnas.95.10.5812.

Salman-Minkov, A., Levi, A., Wolf, S., and Trebitsh, T. (2008). ACC synthase genes are polymorphic in watermelon (*Citrullus* spp.) and differentially expressed in flowers and in response to auxin and gibberellin. *Plant Cell Physiol.* 49, 740–50. doi:10.1093/pcp/pcn045.

Sarin, S., Bertrand, V., Bigelow, H., Boyanov, A., Doitsidou, M., Poole, R. J., et al. (2010). Analysis of multiple ethyl methanesulfonate-mutagenized *Caenorhabditis elegans* strains by whole-genome sequencing. *Genetics* 185, 417–430. doi:10.1534/genetics.110.116319.

Sato, T., and Theologis, A. (1989). Cloning the mRNA encoding 1-aminocyclopropane-1-carboxylate synthase, the key enzyme for ethylene biosynthesis in plants. *Proc. Natl. Acad. Sci. USA.* 86, 6621–5. doi:10.1073/pnas.86.17.6621.

Schaller, G. E., Bleecker, A. B., Hall, A., Binder, B., Schaller, G., and Bleecker, A. (1995). Ethylene-binding sites generated in yeast expressing the Arabidopsis ETR1 gene. *Science* 270, 1809–1811. doi:10.1126/science.270.5243.1809.

Schneeberger, K. (2014). Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* 15, 662–676. doi:10.1038/nrg3745.

Schott-Verdugo, S., Müller, L., Classen, E., Gohlke, H., and Groth, G. (2019). Structural model of the ETR1 ethylene receptor transmembrane sensor domain. *Sci. Rep.* 9, 8869. doi:10.1038/s41598-019-45189-w.

Sega, G. A. (1984). A review of the genetic effects of ethyl methanesulfonate. *Mutat. Res. Genet. Toxicol.* 134, 113–142. doi:10.1016/0165-1110(84)90007-1.

Shiber, A., Gaur, R. K., Rimon-Knopf, R., Zelcer, A., Trebitsh, T., and Pitrat, M. (2008). The origin and mode of function of the Female locus in cucumber. Proceedings of the IXth EUCARPIA meeting on genetics and breeding of Cucurbitaceae, Avignon (France), May 21-24th, 2008, pp. 263-270.

Shifriss, O., and George, W. L. (1964). Sensitivity of female inbreds of *Cucumis sativus* to sex reversion by gibberellin. *Science* 143, 1452–3. doi:10.1126/science.143.3613.1452.

Shinozaki, Y., Nicolas, P., Fernandez-Pozo, N., Ma, Q., Evanich, D. J., Shi, Y., et al. (2018). High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. *Nat. Commun.* 9, 364. doi:10.1038/s41467-017-02782-9.

Shiomi, S., Yamamoto, M., Ono, T., Kakiuchi, K., Nakamoto, J., Nakatsuka, A., et al. (1998). cDNA cloning of ACC synthase and ACC oxidase genes in cucumber fruit and their differential expression by wounding and auxin. *J. Japanese Soc. Hortic. Sci.* 67, 685–692. doi:10.2503/jjshs.67.685.

Shirasawa, K., Hirakawa, H., Nunome, T., Tabata, S., and Isobe, S. (2016). Genome-wide survey of artificial mutations induced by ethyl methanesulfonate and gamma rays in tomato. *Plant Biotechnol. J.* 14, 51–60. doi:10.1111/pbi.12348.

Shnaider, Y., Mitra, D., Miller, G., Baniel, A., Doniger, T., Kuhalskaya, A., et al. (2018). Cucumber ovaries inhibited by dominant fruit express a dynamic developmental program, distinct from

either senescence-determined or fruit-setting ovaries. *Plant J.* 96, 651–669. doi:10.1111/tpj.14051.

Slade, A. J., Fuerstenberg, S. I., Loeffler, D., Steine, M. N., and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.* 23, 75–81. doi:10.1038/nbt1043.

Slade, A. J., and Knauf, V. C. (2005). TILLING moves beyond functional genomics into crop improvement. *Transgenic Res.* 14, 109–115. doi:10.1007/s11248-005-2770-x.

Solano, R., Stepanova, A., Chao, Q., and Ecker, J. R. (1998). Nuclear events in ethylene signaling: A transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* 12, 3703–3714. doi:10.1101/gad.12.23.3703.

Stepanova, A. N., and Alonso, J. M. (2009). Ethylene signaling and response: where different regulatory modules meet. *Curr. Opin. Plant Biol.* 12, 548–555. doi:10.1016/j.pbi.2009.07.009.

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Mol. Plant* 10, 1293–1306. doi:10.1016/j.molp.2017.09.003.

Sun, J. J., Li, F., Li, X., Liu, X. C., Rao, G. Y., Luo, J. C., et al. (2010). Why is ethylene involved in selective promotion of female flower development in cucumber? *Plant Signal. Behav.* 5, 1052–1056. doi:10.4161/psb.5.8.12411.

Switzenberg, J. A., Beaudry, R. M., and Grumet, R. (2015). Effect of *CRC::etr1-1* transgene expression on ethylene production, sex expression, fruit set and fruit ripening in transgenic melon (*Cucumis melo* L.). *Transgenic Res.* 24, 497–507. doi:10.1007/s11248-014-9853-5.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi:10.1111/tpj.12105.

Tao, Q., Niu, H., Wang, Z., Zhang, W., Wang, H., Wang, S., et al. (2018). Ethylene responsive factor *ERF110* mediates ethylene-regulated transcription of a sex determination-related orthologous gene in two *Cucumis* species. *J. Exp. Bot.* 69, 2953–2965. doi:10.1093/jxb/ery128.

Tieman, D. M., Taylor, M. G., Ciardi, J. A., and Klee, H. J. (2000). The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. *Proc. Natl. Acad. Sci. USA.* 97, 5663–5668. doi:10.1073/pnas.090550597.

Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., et al. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.* 7, 19. doi:10.1186/1471-2229-7-19.

Till, B. J., Reynolds, S. H., Weil, C., Springer, N., Burtner, C., Young, K., et al. (2004). Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol.* 4, 12. doi:10.1186/1471-2229-4-12.

Trebitsh, T., Rudich, J., and Riov, J. (1987). Auxin, biosynthesis of ethylene and sex expression in cucumber (*Cucumis sativus*). *Plant Growth Regul.* 5, 105–113. doi:10.1007/BF00024738.

Trebitsh, T., Staub, J. E., and O'Neill, S. D. (1997). Identification of a 1-aminocyclopropane-1-carboxylic acid synthase gene linked to the female (F) locus that enhances female sex expression in cucumber. *Plant Physiol.* 113, 987–995. doi:10.1104/pp.113.3.987.

Tsuda, M., Kaga, A., Anai, T., Shimizu, T., Sayama, T., Takagi, K., et al. (2015). Construction of a

high-density mutant library in soybean and development of a mutant retrieval method using amplicon sequencing. *BMC Genomics* 16, 1014. doi:10.1186/s12864-015-2079-y.

Van der Straeten, D., Van Wiemeersch, L., Goodman, H. M., and Van Montagu, M. (1990). Cloning and sequence of two different cDNAs encoding 1-aminocyclopropane-1-carboxylate synthase in tomato. *Proc. Natl. Acad. Sci. USA.* 87, 4859–63. doi:10.1073/pnas.87.12.4859.

Vicente-Dolera, N., Troadec, C., Moya, M., Del Rio-Celestino, M., Pomares-Viciana, T., Bendahmane, A., et al. (2014). First TILLING platform in *Cucurbita pepo*: A new mutant resource for gene function and crop improvement. *PLoS One* 9. doi:10.1371/journal.pone.0112743.

Vogel, J. P., Schuerman, P., Woeste, K., Brandstatter, I., and Kieber, J. J. (1998). Isolation and characterization of Arabidopsis mutants defective in the induction of ethylene biosynthesis by cytokinin. *Genetics* 149, 417–427.

Wang, C., Xin, M., Zhou, X., Liu, C., Li, S., Liu, D., et al. (2017). The novel ethylene-responsive factor *CsERF025* affects the development of fruit bending in cucumber. *Plant Mol. Biol.* 95, 519–531. doi:10.1007/s11103-017-0671-z.

Wang, D. H., Li, F., Duan, Q.-H., Han, T., Xu, Z.-H., and Bai, S.-N. (2010a). Ethylene perception is involved in female cucumber flower development. *Plant J.* 61, 862–872. doi:10.1111/j.1365-313X.2009.04114.x.

Wang, K., Li, M., and Hakonarson, H. (2010b). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164–e164. doi:10.1093/nar/gkq603.

Wang, N., Wang, Y., Tian, F., King, G. J., Zhang, C., Long, Y., et al. (2008). A functional genomics resource for *Brassica napus* : development of an EMS mutagenized population and discovery of *FAE1* point mutations by TILLING. *New Phytol.* 180, 751–765. doi:10.1111/j.1469-8137.2008.02619.x.

Wang, W., Esch, J. J., Shiu, S.-H., Agula, H., Binder, B. M., Chang, C., et al. (2006). Identification of important regions for ethylene binding and signaling in the transmembrane domain of the ETR1 ethylene receptor of Arabidopsis. *Plant Cell* 18, 3429–42. doi:10.1105/tpc.106.044537.

Weeden, N. F. (1984). Isozyme studies indicate that the genus *Cucurbita* is an ancient tetraploid.. *Cucurbit Genet. Coop. Rep* 7, 84–85.

Wen, X., Zhang, C., Ji, Y., Zhao, Q., He, W., An, F., et al. (2012). Activation of ethylene signaling is mediated by nuclear translocation of the cleaved EIN2 carboxyl terminus. *Cell Res.* 22, 1613–1616. doi:10.1038/cr.2012.145.

Whitaker, T. W., and Davis, G. N. (1962). Cucurbits. Botany, cultivation, and utilization. *Cucurbits. Bot. Cultiv. Util.*

Wien, H. C., Stapleton, S. C., Maynard, D. N., McClurg, C., and Riggs, D. (2004). Flowering, sex expression, and fruiting of pumpkin (*Cucurbita* sp.) cultivars under various temperatures in greenhouse and distant field trials. *HortScience* 39, 239–242. doi:10.21273/hortsci.39.2.239.

Woeste, K. E., and Kieber, J. J. (2000). A strong loss-of-function mutation in *RAN1* results in constitutive activation of the ethylene response pathway as well as a rosette-lethal phenotype. *Plant Cell* 12, 443–455. doi:10.1105/tpc.12.3.443.

Woeste, K. E., Vogel, J. P., and Kieber, J. J. (1999). Factors regulating ethylene biosynthesis in etiolated *Arabidopsis thaliana* seedlings. *Physiol. Plant.* 105, 478–484. doi:10.1034/j.1399-3054.1999.105312.x.

Wuriyanghan, H., Zhang, B., Cao, W.-H., Ma, B., Lei, G., Liu, Y.-F., et al. (2009). The ethylene receptor ETR2 delays floral transition and affects starch accumulation in rice. *Plant Cell* 21, 1473–94. doi:10.1105/tpc.108.065391.

Xanthopoulou, A., Montero-Pau, J., Mellidou, I., Kissoudis, C., Blanca, J., Picó, B., et al. (2019). Whole-genome resequencing of *Cucurbita pepo* morphotypes to discover genomic variants associated with morphology and horticulturally valuable traits. *Hortic. Res.* 6, 94. doi:10.1038/s41438-019-0176-9.

Xie, F., Liu, Q., and Wen, C.-K. (2006). Receptor signal output mediated by the ETR1 N terminus is primarily subfamily I receptor dependent 1. doi:10.1104/pp.106.082628.

Yamasaki, S., Fujii, N., Matsuura, S., Mizusawa, H., and Takahashi, H. (2001). The *M* locus and ethylene-controlled sex determination in andromonoecious cucumber plants. *Plant Cell Physiol.* 42, 608–619. doi:10.1093/pcp/pce076.

Yamasaki, S., Fujii, N., and Takahashi, H. (2000). The ethylene-regulated expression of *CS-ETR2* and *CS-ERS* genes in cucumber plants and their possible involvement with sex expression in flowers. *Plant Cell Physiol.* 41, 608–616. doi:10.1093/pcp/41.5.608.

Yamasaki, S., Fujii, N., and Takahashi, H. (2003). Characterization of ethylene effects on sex determination in cucumber plants. *Sex. Plant Reprod.* 16, 103–111. doi:10.1007/s00497-003-0183-7.

Yang, S. F., and Hoffman, N. E. (1984). Ethylene biosynthesis and its regulation in higher plants. *Annu. Rev. Plant Physiol.* 35, 155–189. doi:10.1146/annurev.pp.35.060184.001103.

Yin, T., and Quinn, J. A. (1995). Tests of a mechanistic model of one hormone regulating both sexes in *Cucumis sativus* (*Cucurbitaceae*). *Am. J. Bot.* 82, 1537–1546. doi:10.1002/j.1537-2197.1995.tb13856.x.

Zhang, J., Guo, S., Ji, G., Zhao, H., Sun, H., Ren, Y., et al. (2019). A unique chromosome translocation disrupting *ClWIP1* leads to gynoecy in watermelon . *Plant J.* doi:10.1111/tpj.14537.

Zhang, J., Shi, J., Ji, G., Zhang, H., Gong, G., Guo, S., et al. (2017). Modulation of sex expression in four forms of watermelon by gibberellin, ethephone and silver nitrate. *Hortic. Plant J.* 3, 91–100. doi:10.1016/J.HPJ.2017.07.010.

Zhao, X.-C., Qu, X., Mathews, D. E., and Schaller, G. E. (2002). Effect of ethylene pathway mutations upon expression of the ethylene receptor ETR1 from Arabidopsis. *Plant Physiol.* 130, 1983–91. doi:10.1104/pp.011635.

Zheng, Y., Wu, S., Bai, Y., Sun, H., Jiao, C., Guo, S., et al. (2019). Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* 47, 1129. doi:10.1093/nar/gky944.

Zuckerkandl, E., and Pauling, L. (1965). Evolutionary divergence and convergence in proteins. *Evol. Genes Proteins*, 97–166. doi:10.1016/B978-1-4832-2734-4.50017-6.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 August 2018
doi: 10.3389/fpls.2018.01049

Phenomic and Genomic Characterization of a Mutant Platform in *Cucurbita pepo*

*Alicia García[1], Encarni Aguado[1], Genis Parra[2], Susana Manzano[1], Cecilia Martínez[1], Zoraida Megías[1], Gustavo Cebrián[1], Jonathan Romero[1], Sergi Beltrán[2], Dolores Garrido[3] and Manuel Jamilena[1]**

[1] Department of Biology and Geology, Research Centers CIAIMBITAL and CeiA3, University of Almería, Almería, Spain, [2] Centro Nacional de Análisis Genómico, Barcelona, Spain, [3] Departamento de Fisiología Vegetal, Facultad de Ciencias, Universidad de Granada, Granada, Spain

The *Cucurbita pepo* genome comprises 263 Mb and 34,240 gene models organized in 20 different chromosomes. To improve our understanding of gene function we have generated an EMS mutant platform, consisting of 3,751 independent M2 families. The quality of the collection has been evaluated based on phenotyping and whole-genome re-sequencing (WGS) results. The phenotypic evaluation of the whole platform at seedling stage has demonstrated that the rate of variation for easily observable traits is more than 10%. The percentage of families with albino or chlorotic seedlings exceeded 3%, similar or higher to that found in other EMS collections of cucurbit crops. A rapid screening of the library for triple ethylene response in etiolated seedlings allowed the identification of four ethylene-insensitive mutants, that were found to be semidominant (*ein1*, *ein2,* and *ein3*) or dominant (*EIN4*). By evaluating 4 adult plants from 300 independent families more than 28% of apparent mutations were found for vegetative and reproductive traits, including plant vigor, leaf size and shape, sex expression and sex determination, and fruit set and development. Two pools of genomic DNA derived from 20 plants of two mutant families were subjected to WGS by using NGS methodology, estimating the density, spectrum, distribution and impact of EMS induced mutation. The number of EMS mutations in the genomes of families L1 and L2 was 1,704 and 859, respectively, which represents a density of 11.8 and 6 mutations per Mb, respectively. As expected, the predominant EMS induced mutations were C > T and G > A transitions (80.3% in L1, and 61% L2), that were found to be randomly distributed along the 20 chromosomes of *C. pepo*. The mutations were mostly affecting intergenic regions, but 7.9 and 6% of the identified EMS mutations in L1 and L2, respectively, were located in the exome, and 0.4 and 0.2% had a moderate and high putative impact on gene functions. These results provide information regarding the potential use of the obtained mutant platform in the discovery of novel alleles for both functional genomics and *Cucurbita* breeding by using direct- or reverse-genetic approaches.

Keywords: *Cucurbita pepo*, EMS, mutant platform, WGS, high-throughput screening, ethylene

Abbreviations: Ein, ethylene insensitive; EMS, ethyl methanosulfonate; L1, L2, mutant family 1 and mutant family 2; NGS, next generation sequencing; SNP, single nucleotide polymorphism; SNV, single nucleotide variant; TILLING, Targeting-induced local lesions in genomes; WGS, whole genome sequencing.

OPEN ACCESS

Edited by:
Jaime Prohens, Universitat Politècnica de València, Spain

Reviewed by:
Rebecca Grumet, Michigan State University, United States
Antonio Ferrante, Università degli Studi di Milano, Italy
Grzegorz Bartoszewski, Warsaw University of Life Sciences, Poland

****Correspondence:***
Manuel Jamilena
mjamille@ual.es

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *21 March 2018*
Accepted: *28 June 2018*
Published: *03 August 2018*

Citation:
García A, Aguado E, Parra G, Manzano S, Martínez C, Megías Z, Cebrián G, Romero J, Beltrán S, Garrido D and Jamilena M (2018) Phenomic and Genomic Characterization of a Mutant Platform in Cucurbita pepo. Front. Plant Sci. 9:1049. doi: 10.3389/fpls.2018.01049

INTRODUCTION

The genus *Cucurbita* comprises three important worldwide-cultivated crop species: *Cucurbita pepo, C. maxima,* and *C. moschata*. The most important cultivated species of the genus is *C. pepo*, consisting of different morphotypes of summer and winter squashes and gourds with great economic importance such as Zucchini, Pumpkin, Vegetable Marrow, Cocozelle, Scallop, Acorn, Straightneck and Crookneck. In the year 2014 the cultivated area of *C. pepo* reached nearly 2 million of hectares and a production of 25 million of tons (FAOSTAT, 1961).

The genome of *C. pepo* has been recently sequenced and annotated[1] (Montero-Pau et al., 2017). It consists of 263 Mb, a scaffold N50 of 1.8 Mb, and 34,240 gene models, organized in 20 chromosomes that cover about 93% of the assembled sequence. A duplication event has been detected in the genome of *Cucurbita*, happening about 20 Mya ago, after the separation from other genus of the *Cucurbitaceae* family such as *Citrullus* (watermelon) and *Cucumis* (melon and cucumber) (Montero-Pau et al., 2017; Sun et al., 2017).

Although the drafts of the genomes of *C. pepo* and other cucurbit species are becoming more complete, little is known about specific gene functions in these crops. Chemically induced plant mutant platforms, such as those generated by EMS, a chemical mutagen that induces single randomly distributed nucleotide changes in DNA (Sega, 1984; Greene et al., 2003), have become an important source of variability for both functional genomic analyses and for plant breeding programs. Once established, the mutant platforms can be used for direct phenotyping, but also for DNA screenings, allowing the detection of single point mutations in a number of specific genes.

High-throughput screenings based on plant phenotyping of large mutant collections are becoming increasingly important as a source of new useful mutants in crop species. The difficulty of direct screenings lies in the large number of plants to be evaluated for each of the characters under study. Since mutations are mostly recessive, the identification of a mutant phenotype requires 8–10 plants to be phenotyped per family, and therefore the screening of 1,000 mutant families requires the assessment 8,000–10,000 plants. Massive screenings based on seedling early traits have been developed for the detection of agronomic interesting mutants, including mutants with tolerance to biotic and abiotic stresses, as well as insensitive mutants to certain plant hormones or to chemical treatments with specific herbicides or pesticides. For example, the molecular bases of ethylene-regulated processes have been discerned thanks to the analysis of *Arabidopsis* mutants which are altered in the seedling triple response to ethylene (Bleecker et al., 1988). The response of seedlings to ethylene have been successfully used to screen different EMS mutant collections in *Arabidopsis* (Guzmán and Ecker, 1990), which allowed the identification of a number of mutants altered in the ethylene pathway (Roman et al., 1995).

Indirect screenings were become more popular since the development of the TILLING (Targeting Induced Local Lesions in Genomes) reverse genetic approach (Colbert et al., 2001). This methodology was firstly applied to detect allelic variants of an specific gene in *Arabidopsis* EMS mutant collections (Greene et al., 2003), but its utilization was rapidly extended to a number of crop species, including rice (Till et al., 2007), maize (Till et al., 2004), barley (Caldwell et al., 2004), *Brassica napus* (Wang et al., 2008), tomato (Minoia et al., 2010), among others. In cucurbit species, TILLING platforms were developed, allowing the detection of new allelic variants for plant breeding in *C. melo* (Dahmani-Mardas et al., 2010; González et al., 2011) and *C. sativus* (Boualem et al., 2014; Fraenkel et al., 2014). An EMS mutant collection has already been published in *C. pepo*, and different target genes were assessed by TILLING (Vicente-Dólera et al., 2014), but the number of mutant squash families is still very poor.

In this paper we present the development and utilization of a new EMS mutant platform in *C. pepo*, made of 3,751 M2 families. The quality and utility of the collection was confirmed by phenomic screenings of the mutant plants at both seedling and adult stage of development, by using a forward high-throughput screening for ethylene insensitivity, and by assessing the mutation rate and distribution of mutations in two families by whole genome re-sequencing. Results demonstrated that the collection has sufficient quality to address the identification of new *C. pepo* mutants throughout forward genetics, as well as the high density and random distribution of induced mutations in the genome of this crop species for reverse genetics. The established mutant platform constitutes therefore an important resource for functional genomic studies, but also a source of new alleles for squash breeding programs.

MATERIALS AND METHODS

Generation of the EMS Mutant Platform

The mutant library was generated in an inbred line derived from the Spanish landrace MUC16. Before starting the generation of the EMS mutant collection, we established the optimal EMS concentration producing the higher mutation density without much altering the seed germination rate, neither the fertility of the M1 plants. For this end, batches of 200 mature seeds of MUC16 were immersed in bottles containing 0–2% EMS in 200 ml of deionized water and stirred in a shaker for either 12 or 24 h at 22°C in darkness. After the EMS treatment, seeds were washed twice in 200 ml of 3% $Na_2S_2O_3$ buffer for 30 min at room temperature with gentle shaking, followed by three washes in 200 ml distilled water for 30 min each. Control seeds were treated with distilled water in the same manner. Treated seeds were germinated in a commercial nursery following standard local practices. After evaluating the germination rate, seedlings were grown under standard greenhouse conditions in Almería (Spain) to obtain mature M1 plants. Each M1 plant was then self-pollinated, and M2 seeds extracted from individual fruits and stored. The fertility of M1 plants was assessed as the percentage of plants producing more than 20 viable seeds per fruit, i.e., those fruits producing less than 20 seeds were considered to be derived from male- or female-sterile plants. From those experiments we concluded that the optimal

[1] cucurbitgenomics.org

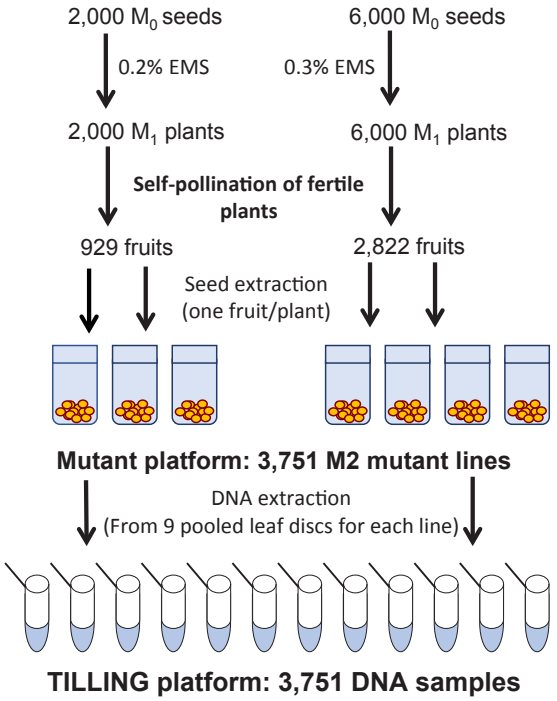


FIGURE 1 | Steps in the construction of the mutant and TILLING platforms of *C. pepo*, each one composed of 3,751 seed and DNA samples, respectively.

mutagenic treatment was that containing 0.3% EMS for a total of 12 h.

To generate the *C. pepo* mutant collection, we treated 6,000 and 2,000 MUC16 seeds with either 0.3 or 0.2% EMS for 12 h, respectively. Those EMS doses did not reduce the germination rate of the MUC16 seeds but lessened the fertility of the plants to nearly 50%. Plants derived from the 6,000 and 2,000 treated seeds were selfed to achieve a total of 2,822 and 929 fruits with more than 20 viable seeds (**Figure 1**). Consequently, the generated mutant collection comprises a total of 3,751 M2 families. Regarding the mutant seeds availability, we currently have an average of 160 seeds per M2 family, and around 600 seeds for each of the 900 M3 families that were already propagated.

Screening of the Mutant Library and Generation of a TILLING Platform

The quality of the mutant collection was firstly tested by evaluating the percentage of easily observable traits in 9 seedlings of each mutant family. A total of 30,008 seedlings were grown, and the alterations in the development and pigmentation of cotyledons and first leaves, in comparison with the background genotype MUC16, recorded for each plant. For each M2 mutant family we also collected together one leaf disk from eight individual seedlings for DNA extraction, developing so a TILLING platform consisting of 3,751 DNA samples.

The M2 plants from 284 families in the 0.3% treatment were also grown to maturity under standard greenhouse conditions, determining the percentage of alterations in vegetative and reproductive developmental traits, including plant vigor and height, leaf morphology, root conformation, female and male flower development, sex related traits, and fruit color, size, shape. These M2 families were also selfed to obtain the M3 generation.

Assay of Triple Response to Ethylene

To test the quality of the squash mutant collection we also performed a high-throughput screening of the 3,751 M2 families for mutants in the ethylene-response pathway. The triple response of etiolated seedlings to ethylene was so used to detect M2 families segregating for Ein mutants. Eight seeds of each M2 family were sown in seedlings trays, germinated for 2 days in the absence of ethylene and then introduced into a growth chamber containing 50 ppm of ethylene in darkness for a total of 5 days. The Ein seedlings were perfectly distinguishable since they did not reduce the length of their hypocotyls, protruding therefore over the rest of the ethylene-sensitive seedlings in the growth chamber. By using this ethylene test we detected four segregating M2 families. The ethylene insensitivity of these four families was confirmed by evaluating the response to ethylene within a higher number of plants in each selected M2 family. The Ein plants were then backcrossed with the inbred line MUC16 for two generations and the resulting progeny (BC_2) selfed to achieve the BC_2S_1 generation. These generation was again evaluated by the ethylene triple response.

Data were analyzed by multiple comparisons by analysis of variance (ANOVA) with significance level $p < 0.05$, and each two means were compared with the Fisher's least significant difference (LSD) method. The chi-square analyses were performed using the statistical software Statgraphics Centurion XVI.

Library Preparation and Sequencing

Whole genome sequencing was performed at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain). Four DNA samples were sequenced for two random M_2 families from the *C. pepo* mutant collection: family 435 corresponding to mutant family 1 (L1), and family 1717 corresponding to mutant family 2 (L2). Each sample comprises 10 random plants from each family. Paired-end multiplex libraries were prepared according to manufacturer's instructions with KAPA Library Preparation kit (Kapa Biosystems). Libraries were loaded to Illumina flowcells for cluster generation prior to producing 126 base read pairs on a HiSeq2000 instrument following the Illumina protocol. Base calling and quality control was done with the Illumina RTA sequence analysis pipeline according to manufacturer's instructions. Based on the genomic sequence of *Cucurbita pepo* (genome v4.1) the average fold effective coverage of all the samples was in the range between 13.18 and 17.90 (**Supplementary Table 1**).

Data Analysis

Sequencing reads were trimmed to avoid any remaining adaptors. Resulting reads were mapped to the *C. pepo* genome v4.1[2] using the GEM3 toolkit (Marco-Sola et al., 2012) allowing up to

[2] cucurbitgenomics.org

8% mismatches. Alignment files (BAM format) containing only properly paired, uniquely mapping, reads were processed using picard tools version 1.110 (Picard Tools - By Broad Institute) to add read groups and remove duplicates. The Genome Analysis Tool Kit (GATK) (McKenna et al., 2010) version 3.6 was used for local realignment and base recalibration. Processed BAM files were submitted to variant calling for SNVs and small insertions and deletions using HaplotypeCaller (GATK v3.6). Functional annotations were added to the resulting VCF using SnpEff (Cingolani et al., 2012) with the gene annotation provided for *C. pepo* genome v4.1[2].

Variant Filtering

Variants were removed when depth of coverage was less than 8 or greater than 100 in at least one of the samples of each family (L1 and L2). This allowed us not only to exclude positions for which the coverage depth was low, but also positions that might fall in segmental duplications. From the whole genome (260 Mb), after filtering we keep 174 Mb for family L1 and 178 Mb for family L2.

Wrongly mapped reads are difficult to account for and can result in an increased false discovery of genetic variants. In order to remove undetected duplications or repetitive regions that could mislead the alignment program, we further filtered out genomic regions with unusually high number of variants. To identify those regions, a sliding window approach was used, with a window size of 1 kb and a step size of 500 bp. All genomic 1 kb windows with more than one variant where discarded. As a result, we discarded 25 Mb of total genomic regions from L1 and 26 Mb from L2.

To determine if a mutation has been caused by EMS in one of the families, it is necessary to have reliable information for that particular position in both families (L1 and L2). After applying the coverage filter removing the hyperpolymorphic regions the intersection of the L1 and L2 genomic positions was computed. The shared reliable regions contained a total of 144.5 Mb.

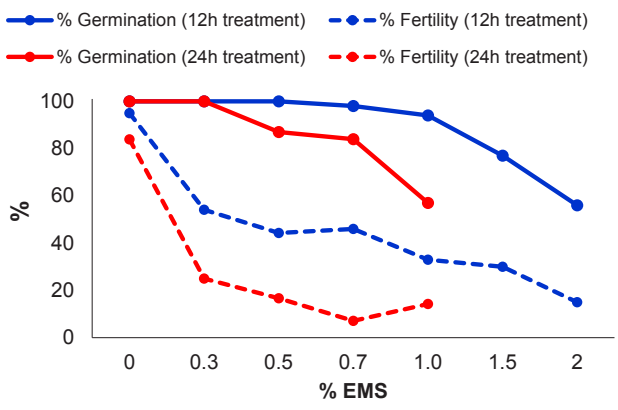


FIGURE 2 | Effect of different EMS treatments on M1 seed germination and plant fertility.

RESULTS

Generation of an EMS Mutant Collection in *C. pepo*

In order to optimize the EMS induced mutagenesis, batches of 200 seeds of MUC16, a *C. pepo* inbred line that was also used for genome sequencing (Montero-Pau et al., 2017), were treated with different concentrations of EMS for 12 and 24 h. The efficiency of the mutagenesis treatments was then estimated by assessing two parameters on M1 plants: seed germination and plant fertility. **Figure 2** shows the reduction in the percentage of germination and in the fertility of M1 plants subjected to the different treatments.

In seeds treated for 12 h, no significant reduction in germination rates was observed in 0.3–1% EMS treatments, but in 1.5 and 2% EMS treatments the germination was highly reduced (**Figure 2**). Most of the M1 plants which were treated for 12 h, exhibited growth retardation at seedling stages, but all of them recovered and flowered. We found an inverse

FIGURE 3 | Seedling of *C. pepo* M2 mutant families, showing certain phenotypic alterations: three-cotyledon seedlings, deformation in cotyledons and leaves, variegation and depigmentation in cotyledons and leaves, albino seedlings and dwarfisms.

TABLE 1 | Percentage of phenotypic variations in the seedlings of the 3,751 M2 mutant families of *C. pepo*.

	0.2% EMS (929 families)		0.3% EMS (2,822 families)		Total (3,751 families)
	Number of families	Frequency (%)	Number of families	Frequency (%)	Frequency (%)
Albino	5	0.54	33	1.2	1.04
Tri-cotyledon	13	1.4	29	1.03	1.12
Negative gravitropism	3	0.32	36	1.28	1.04
Cotyledon deformation	10	1.08	95	3.4	2.83
Dwarfism	9	0.97	62	2.2	1.89
Variegation	4	0.43	33	1.2	1.01
Depigmentation	18	1.94	50	1.77	1.81
Orange color	0	0	2	0.07	0.05
Insensitive-ethylene	0	0	4	0.14	0.11
Total M2 families	**62**	**6.67%**	**244**	**12.19%**	**10.82%**

correlation between mutagen dose and M1 plant fertility, assessed as the number of selfed fruits, each one having more than 20 viable seeds (**Figure 2**). At 1.5 and 2% EMS, 70 and 85% of the plants were considered sterile because they either did not set fruit or yielded fruits with less than 20 seeds after selfing.

In EMS treatments for 24 h, the 1.5 and 2% treatment were omitted. Even so, the germination rate was highly reduced,

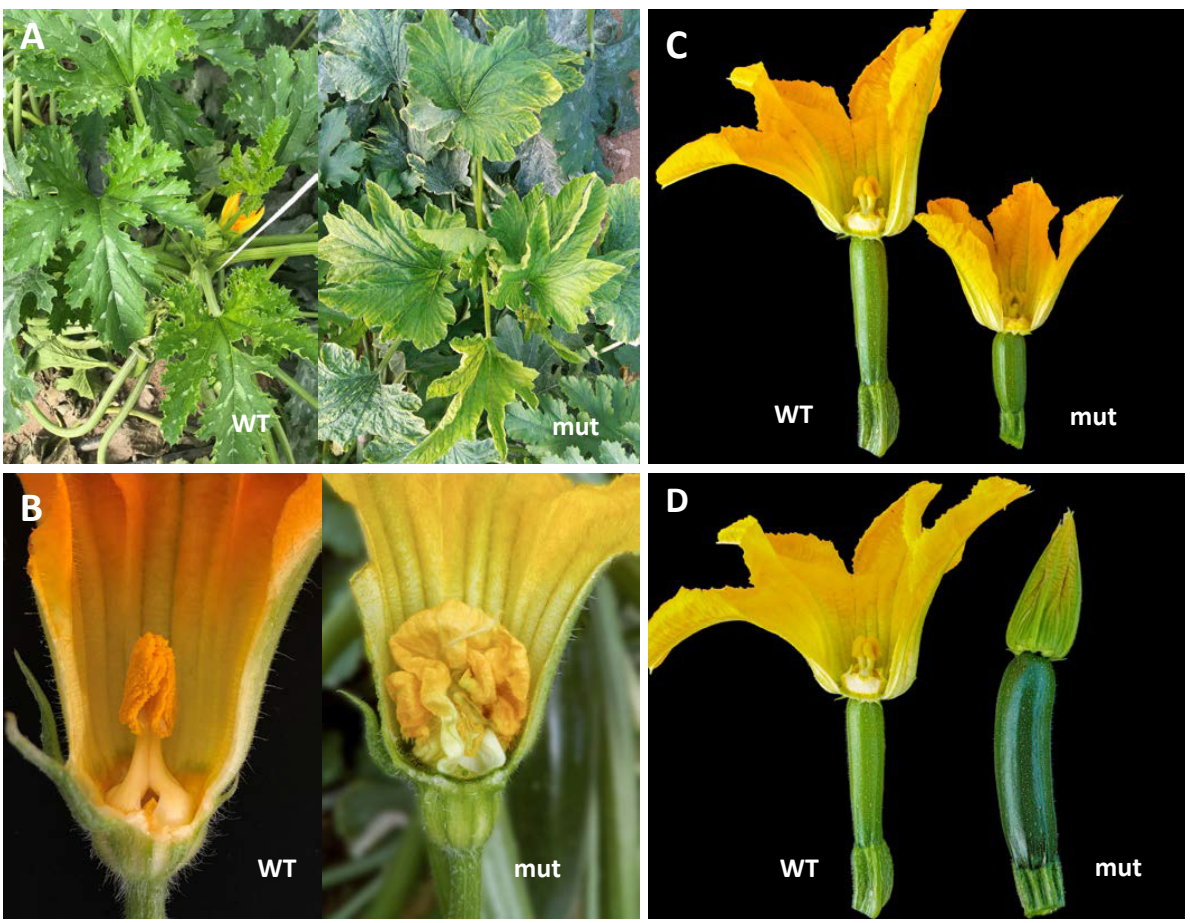


FIGURE 4 | Mutant phenotypes observed in M2 mutant families. **(A)** Alterations in leaf morphology. **(B)** Alterations in male flower development. **(C)** Alterations in female flower size, and ovary shape and size. **(D)** Alterations in fruit set and female flower development (the mutant fruit has a parthenocarpic development).

TABLE 2 | Percentage of phenotypic variations in adult plants from 284 M2 families derived from the 0.3% EMS treatment.

Phenotypic variation	Number of families	Percentage (%)
Vegetative development	26	9.15
Male flower	14	4.93
Female flowers	20	7.04
Fruit set and development	20	7.04
Total families affected	**80**	**28.17%**

reaching 57% germination after 1% EMS treatment (**Figure 2**). The fertility of the M1 plants was greatly reduced in EMS treatments for 24 h, showing less than 75% fertility in the treatments with the different doses (**Figure 2**). Taken together, the results demonstrated that, even at levels that barely reduced the germination rate, the fertility of plants was strongly affected. Therefore, the most appropriated EMS dosage to produce the mutant collection was selected based on M1 plant fertility. The 0.3% EMS for 12 h was selected as the most suitable treatment for mutagenesis of MUC16 seeds. This EMS dosage did not affect the germination rate of the M1 plants, but reduced M1 plant fertility to about 54% (**Figure 2**). Fertile plants yielded fruits that contained more than 100 seeds after selfing, therefore not compromising the production of the M2 generation. In addition, we selected a slightly reduced EMS concentration (0.2%) to obtain one third of the mutant families in our collection.

For generating the final mutant collection, 8,000 MUC16 seeds were mutagenized; 6.000 were treated with 0.3% EMS and 2000 seeds with 0.2% EMS for 12 h at 22°C. The germinated seed was grown in a greenhouse and each M1 fertile plant selfed to obtain the M2 offspring. A total of 3,751 M2 families were obtained, 929 and 2,822 for the 0.2 and 0.3% doses of EMS, respectively (**Figure 1**). Nine plants were grown from each mutagenized family, and leaf material from each plant was collected and mixed for DNA extraction. A total of 3,751 samples of DNA were isolated to establish a TILLING platform of the *C. pepo* mutant collection (**Figure 1**).

Phenotypic Variations in the Mutant Collection

The quality of the mutant collection was firstly tested by evaluating the phenotypes of 9 plants of each M2 family at the seedling stage of development, focusing on morphological traits that were easily observable in seedlings. The observable changes, which appeared in the cotyledons and leaves of some plants in relation to the genetic background MUC16, were recorded for each M2 family. **Figure 3** shows most of the seedling mutant phenotypes in cotyledons and leaves. A number of M2 families showed alterations in the number and development of cotyledons, other in the development of leaves, and others in the pigmentation of vegetative organs. The percentage of families with typical mutations such as albino plants was 0.54 and 1.2% for the families derived from either 0.2 or 0.3% EMS treatments (**Table 1**). Other mutant phenotypes, including deformations in cotyledons and leaves, were at higher frequency than albinisms, and the frequency of mutant families with depigmentation in cotyledons and leaves was even higher in those derived from the 0.2% EMS treatment, indicating that density of mutations in the plants derived from this weaker treatment is also elevated (**Table 1**). Therefore, at the seedling stage, the total number of phenotypic variations in the mutant library was estimated to be 10.82% (**Table 1**).

In adult plants the estimation of phenotypic variations was increased, since we not only phenotyped the vegetative organs of the plant but also flowers and fruits (**Figure 4**). From a total of 284 M2 families, vegetative development alterations were observed

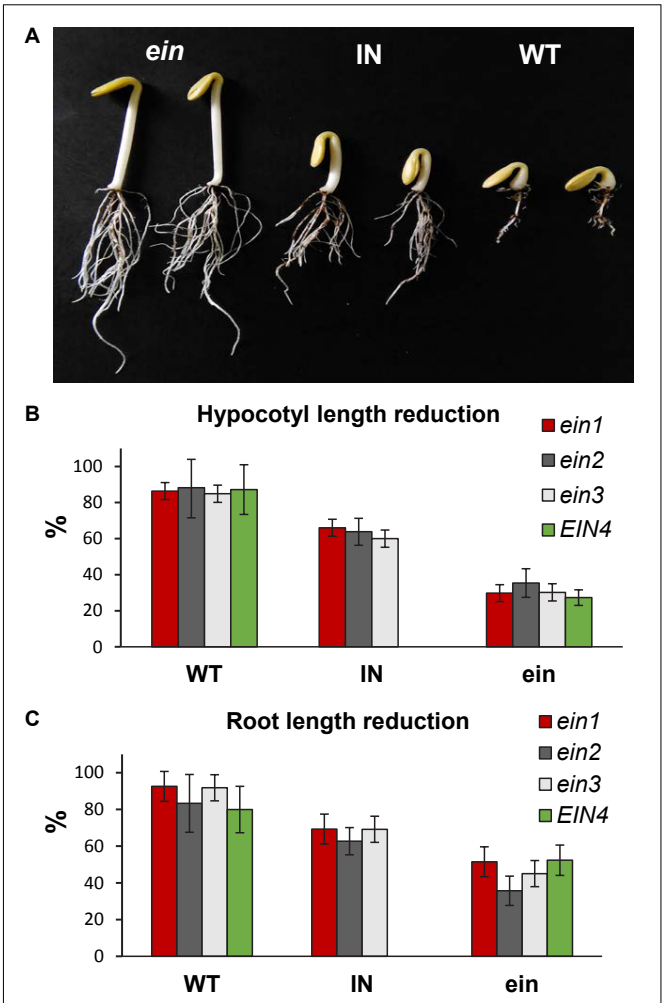


FIGURE 5 | (A) Ethylene triple response phenotypes in etiolated seedlings of the *ein1* mutant family of *C. pepo*. WT plants showed a positive triple response, characterized by a reduction in the length of hypocotyl and root, and a thicker hypocotyl. The ethylene insensitive (*ein*) plants do not respond to ethylene and show longer and thinner hypocotyl, and longer roots. Certain plants in the family showed an intermediate (IN) triple response phenotype. **(B,C)** Percentage of reduction of hypocotyl and root length of ethylene treated WT and *ein* mutants in comparison with untreated seedlings of the background genotype MUC16. Errors bars represent standard errors. Note that in the *ein1-3* mutant families there is a distinguishable intermediate (IN) phenotype to ethylene response.

TABLE 3 | Segregation of ethylene insensitive mutants in the backcrossing (BC_1 and BC_2), and the first selfed (BC_2S_1) generations.

Mutant family	Generation	Number of plants				χ^2	*p*-value
		WT	INT	*ein*	Expected segregation		
ein1	BC_1	–	62	–		–	–
	BC_2	20	24	–	1:1	0.36	0.54
	BC_2S_1	116	187	115	1:2:1	4.64	0.10
ein2	BC_1	–	79	–		–	–
	BC_2	42	41	–	1:1	0.01	0.91
	BC_2S_1	129	237	119	1:2:1	0.66	0.72
ein3	BC_1	–	42	–		–	–
	BC_2	31	27	–	1:1	0.28	0.60
	BC_2S_1	98	178	86	1:2:1	0.89	0.64
EIN4	BC_1	34	–	30	1:1	0.25	0.62
	BC_2	120	–	125	1:1	0.10	0.75

Plants were classified according to their triple response to ethylene: WT, ethylene sensitive seedlings; INT, intermediate response to ethylene; ein, ethylene insensitive seedlings. EIN4 plants only produce male flowers and were unable to be selfed.

in 9.15% of the mutant families (**Table 2**), a similar frequency of vegetative development variations detected in the complete library at the seedling stage. However, in adult plants we also detected different alterations in the morphology of female and male flowers (7.04 and 4.93%, respectively) as well as in fruit color, size and shape (7.04%). The total frequency of phenotypic variations was increased to 28.17% which represents a high frequency of variation for forward genetic approach.

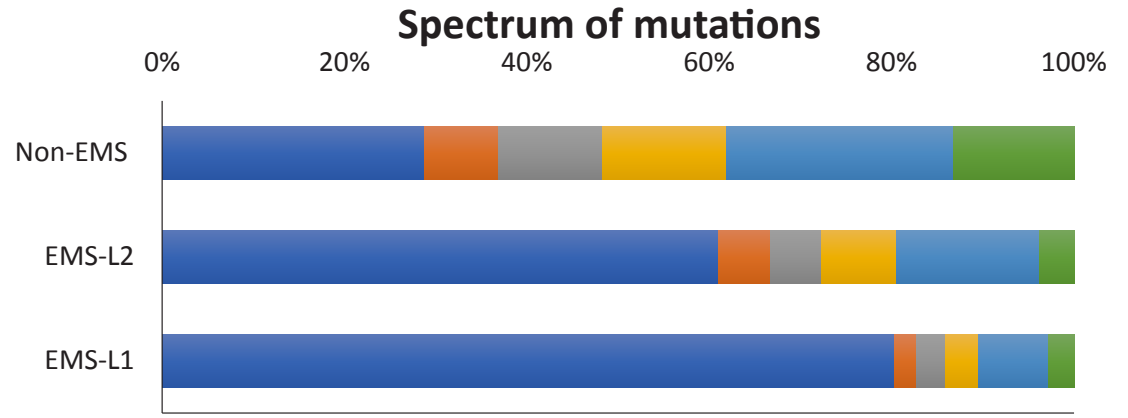
High-Throughput Screening of the Collection for Ethylene Insensitivity

To test the quality of the *C. pepo* mutant library for forward genetic analysis, we performed a high-throughput screening for ethylene insensitivity by using the triple response of etiolated seedlings to ethylene. Nine seeds of each family were germinated and treated with gaseous ethylene for 48 h. In the WT genotypic background (line MUC16) as well as in the vast majority of the M2 families, all plants showed a positive triple response to ethylene. Seedlings displayed a decrease in the length of hypocotyl and roots, an increased hypocotyl thickness and greater curvature of the apical hook, characteristic that were rapidly observed in the germinating seeds (**Figure 5**). Four out of 3,751 families showed a reduced response to ethylene treatment, indicating that some of the plants were insensitive to this hormone. The identified mutants, designated *ein1*, *ein2*, *ein3* and *EIN4*, developed larger and thinner hypocotyls, as well as lager roots in comparison with WT plants of the same family (**Figure 5**). In the *ein1*, *ein2* and *ein3* families, some of plants showed an intermediate phenotype between WT and *ein* mutants.

Only 1–4 plants out of nine analyzed seeds in each mutant family showed an Ein phenotype. To confirm the ethylene insensitivity of these mutant families, we increased the number of plants in segregating progenies of each family and reduced the number of other mutations in each of the families by backcross. The M_2 mutant plants were backcrossed with the background WT genotype MUC16 for two generations, and the BC_2 progeny selfed to obtain the BC_2S_1 generation. The backcrossed plants had the same Ein phenotype as the original M2 plants. Segregations also demonstrated that the four mutations were semidominant, and that intermediate phenotype for ethylene triple response corresponded to heterozygous plants for the mutations (**Table 3**). Furthermore, by using a higher number of plants for each mutant family, we assessed the reduction in the length of hypocotyls and roots of ethylene treated seeds in comparison with untreated seeds of the background genotype MUC16 (**Figure 5**). In the segregating progenies of *ein1*, *ein2* and *ein3* mutants, we distinguished three phenotypes, although only two phenotypes were found in the *EIN4* family (**Figure 5**).

Density and Spectrum of EMS Mutations

Next generation sequencing technology makes whole-genome sequencing a practical method for mutation identification and mapping. DNA from 40 plants in segregating progenies of two mutant families (L1 and L2) was used for WGS. Since WGS was made simultaneously with the phenotyping of seedlings and plants, these two families were randomly selected from the collection. They were later phenotyped, but no alterations were detected in easily observable morphological traits. Four samples, each one containing a pooled DNA from at least 10 plants in segregating progenies, were sequenced for each family. Sequencing data from the eight samples showed a mean coverage of 92.4 and 95.5% of the whole genome for L1 and L2, respectively. After variant calling for each individual sample, SNVs with GQ > 20 were firstly filtered for a minimum coverage threshold of 8 reads in at least one sample per family, and maximum coverage threshold of 100 reads (median coverage per sample was 35 reads). The regions with higher coverage were filtered out because their high number of SNVs is most likely caused by wrong mapped interspersed DNA sequences. This filtered out 15% of the bases, resulting in a genome coverage of 67 and 67.8% for families L1 and L2, respectively (**Supplementary Table 4**).



	EMS-L1	EMS-L2	Non-EMS
GC>AT	1368	524	1158
GC>CG	41	49	331
GC>TA	54	48	457
TA>AT	62	71	551
TA>CG	132	134	1001
TA>GC	47	33	533

FIGURE 6 | Spectrum and frequency of and EMS and non-EMS mutations in L1 and L2 families. EMS mutations correspond to those that were heterozygous and specific for one of the mutant family. Homozygous family specific SNVs and shared SNVs between L1 and L2 families were considered non-EMS.

A second filtering step was based on the density of variants, filtering out those SNVs that had another variant closer than 1,000 bp because they were likely caused by sequencing or mapping errors. That reduced the coverage of the reference genome to about 60% in the two families, but also reduced the number of variants to 6,140 in L1 and 5,252 in L2. Most of these variants (4,031) were common to both L1 and L2 (**Supplementary Table 4**), likely representing spontaneous nucleotide polymorphisms in MUC16 genetic background. They account for 65.7 and 76.8% of SNVs in L1 and L2, respectively. Those that were homozygous in both L1 and L2 (3,136) were probably fixed during the 2–3 extra rounds of selfing that MUC16 reference genotype was subjected before EMS treatments. The other shared SNVs could be spontaneous polymorphisms still segregating in one or the two families. Given that the MUC16 line has more than nine generations of selfing and taking into account that *C. pepo* genome is duplicated (Montero-Pau et al., 2017), we cannot rule out that these shared SNVs were the result of mapping errors in reads coming from duplicated genomic regions.

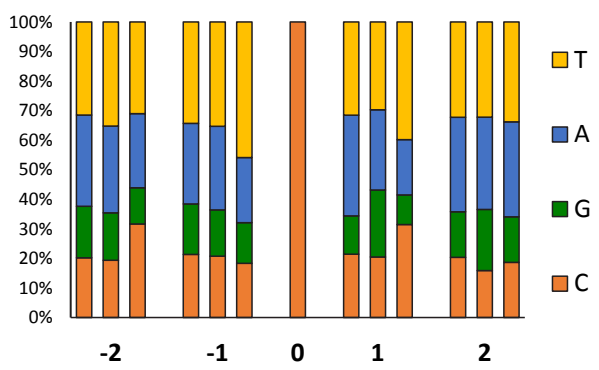


FIGURE 7 | Frequency of the sequences flanking EMS induced mutations. At each position (–2, –1, +1 and +2), left bars correspond to the frequency of bases flanking C in the background genome, the middle bars, to flanking sites of the non-EMS CG > TA variants (natural polymorphisms shared by the two families respect to the reference genome), and the right bars to the canonical EMS mutations (CG > AT).

The specific variants sum a total of 2,209 in L1 and 1,226 in L2 from a total of 144.5 Mb (55.4% of genome) shared by the two families after filtering (**Supplementary Table 4**). Given the high percentage of spontaneous mutations among SNVs, the homozygous family specific SNVs would mostly be spontaneous polymorphisms that were fixed for the alternative allele in each one of the families after mutagenesis. Therefore, we discarded these family specific homozygous SNVs as truly EMS mutations.

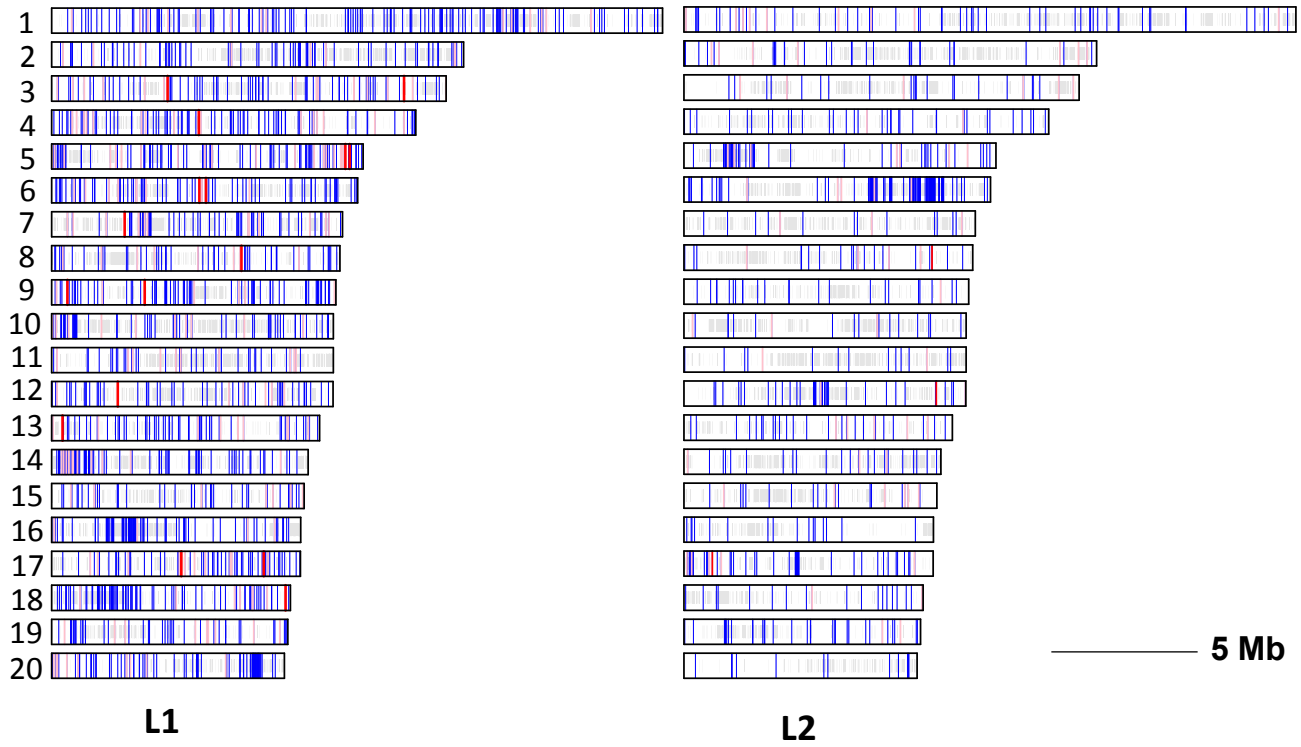


FIGURE 8 | Chromosome distribution (1–20) of EMS-induced mutations in L1 and L2 mutant families. Depending on the predicted impact, EMS-induced variants are plotted as red (high impact), pink (moderate) or blue (low or very low impact). Regions discarded because of the high density of variants are shaded in gray.

The family specific heterozygous SNVs (1,704 and 859 in L1 and L2, respectively) were therefore selected as the tangible EMS induced DNA mutations, which reveals a mutation density of 11.8 mutation/Mb (1 mutation/85 kb) in L1, and 6 mutations/Mb (1 mutation/167 kb) in L2.

The most frequent mutations were the EMS canonical transitions GC to AT (80.3% in L1 and 61% in L2), followed by TA to CG transitions (7 and 15.6%) (**Figure 6**). The frequency of transversion (GC > CG, GC > TA, TA > AT and TA > GC) was relatively low, and summarizes 12.7% in L1 and 23.4% in L2 (**Figure 6**). The transition to transversion ratio was therefore 6.9 and 3.3% in L1 and L2, respectively. The spectrum of mutations in the fraction of SNVs that were discarded as non-EMS (those that were shared by the two families, and those that were homozygous in only one of the families), was not enriched in GC to AT transitions (**Figure 6**), suggesting that they were not likely induced by the mutagen but were polymorphism of the background genotype.

To investigate sequence biases in EMS-induced mutations, we analyzed the frequency of preferential bases flanking ± 2 positions around the most frequent transitions (GC to AT) respect to the genome background (**Figure 7**). In the background genome, the two bases surrounding a C, were more frequently A and T than C and G. In −1 and +1 positions of the most numerous GC to AT transitions, pyrimidines are more frequent than expected, with significant biases to T in the −1 position and to T and C in the +1 position. No bias was observed in the +2 position, but an excess of Cs occurred at the −2. In comparison with the background genome, these nucleotide biases were not detected around the GC to AT spontaneous SNVs observed in the mutant families (**Figure 7**).

Distribution and Impact of EMS Mutations

The distribution of EMS mutations among the 20 chromosomes of *C. pepo* is shown in **Figure 8**. The mutations were evenly distributed throughout the genome with a density ranging from 5 to 11.7 mutations/Mb in chromosomes 15 and 14 of L1, and 1.5 to 13.6 mutations/Mb on chromosomes 7 and 6 of L2. No preferential distribution was detected among euchromatic and heterochromatic regions in the different chromosomes.

The causative effects of the mutation on gene function was studied by using the SnpEff program. On the base of *C. pepo* genome annotation, many of the EMS mutations were located in regions with low predicted impact on gene function, including intergenic regions (23.6%), downstream and upstream (55%) regions and introns (11.1%). A reduced number of EMS mutations affected, however, to exons (7.3%), 5′- and 3′-UTR regions (2.4%) and splice sites (0.5%), which could likely alter gene function (**Table 4**). Of the 422 exonic mutations 281 (66.6%) were missense mutations, 11 (2.6%) nonsense mutations, and 130 (30.8%) silent (**Table 4**). Regarding their impact, 0.3% of the detected EMS mutations had a predicted high- (e.g., nonsense, frameshift, and splice acceptor or donor mutations), 4.8% a

TABLE 4 | Location and functional impact of EMS mutations in L1 and L2 mutant families of *C. pepo*.

	Mutation	L1	L2	Total	%
Gene region	Exon	311	111	422	7.3
	Intron	479	166	645	11.1
	Splice sites	23	6	29	0.5
	Intergenic	821	548	1,369	23.6
	5′- and 3′- UTR	101	41	142	2.4
	Downstream and upstream	2,202	990	3,192	55
Protein change	Missense	207	74	281	66.6
	Nonsense	9	2	11	2.6
	Silent	95	35	130	30.8
Impact	Low-impact mutations	117	40	157	3.5
	Moderate-impact mutations	207	74	281	6.3
	Modifier-impact mutations	2,776	1,197	3,973	89.7
	High-impact mutations	16	3	19	0.4

*Predicted low-impact mutations correspond to synonymous and splice region variants; Moderate-impact mutations, to missense variants, and high-impact mutations, to stop gained, and splice acceptor and splice donor variants (**Supplementary Tables 2, 3**).*

moderate- (e.g., missense mutations), 92.1% a modifier- (e.g., intron and intergenic mutations), and 2.8% a low-impact (e.g., synonymous mutations) on gene function.

Of the 19 high-impact mutations, 10 were nonsense, and 3 and 5 occurred on splice acceptor and splice donor sites of target genes, respectively (**Table 4**). The genes affected by a premature stop codon, which are likely those with higher impact on gene function, code for proteins like: E3 ubiquitin-ligase PRT6 and probable RNA helicase SDE3 in family L2, and for chloroplastic glyceraldehyde-3-phosphate dehydrogenase; multi-bridging factor 1a-like; probable LRR receptor like serine threonine- kinase At4g26540; low-temperature-induced cysteinase; nuclear transcription factor Y subunit B-9; ADP-ribosylation factor-related 1; Protease 2, and aquaporin TIP2-1 in family L1 (**Supplementary Tables 2, 3**).

DISCUSSION

Quality and Usefulness of the *C. pepo* EMS Mutant Library

A mutant platform has been established in the crop species *C. pepo*, consisting of 3,751 M2 families with an average mutation density of 8.9 mutations/Mb (1 mutation per 112 kb). The artificial mutagenesis was found to generate more than 10% of morphological variants in early seedling developmental stages, and more than 30% of morphological alterations in adult plants. This mutation density, detected by both phenomic and genomic approaches, is quite similar or higher to that observed in other plant EMS collections (Dahmani-Mardas et al., 2010; González et al., 2011; Boualem et al., 2014; Vicente-Dólera et al., 2014), indicating that the library will be a high-quality tool for both forward and reverse genetic analyses.

The efficiency of chemical mutagenesis is variable, depending not only on the nature and dosage of the mutagen, but also on the considered species and the background genotype (Chen et al., 2014). The *C. pepo* library was generated with a lower EMS dosage than in other plant species, including the related cucurbit species. *C. melo* collection was generated by 1.5% EMS, while that of *C. sativus* was generated by using 0.5 and 0.75% EMS (Dahmani-Mardas et al., 2010; Boualem et al., 2014). Similar doses (0.7–1.5%) were used in tomato (Minoia et al., 2010), maize (Till et al., 2004) and rice (Till et al., 2007). Despite the use of lower EMS dosage, the *C. pepo* library did not reduce the mutation density, suggesting that *C. pepo* is more sensitive to EMS than other species. Seed germination was similarly affected in the different species, but the more harmful effect of EMS treatment in squash was observed on plant fertility. Consequently, the selection of EMS dose could not be made on the basis of seed germination but based on M1 plant fertility. At the end an EMS dosage was utilized (0.2 and 0.3%) that resulted in 100% of M1 seed survival, but reduced fertility to about 50%, and allowed therefore the production of large populations of viable M2 seeds.

The alterations found in seedlings morphological traits, which is a good indicator of a high-quality mutant collection (Wang et al., 2008), reached 10.82%, is similar to that reported for other well characterized mutant collections. Thus, the percentage of M2 families with albino and chlorophyll deficient families in our collection (1.85%) was very similar to the 1.3% rate found in a previous squash library (Vicente-Dólera et al., 2014), as well as the 0.6 and 2.1% rates observed in the mutant collection of *C. sativus* (Boualem et al., 2014), and muskmelon (González et al., 2011), respectively. Similar frequency of albino mutants, were found in tomato (1.7%; Minoia et al., 2010) and in common bean (1.53%; Porch et al., 2009). The frequency of the dwarf phenotype in our library (1.89%) was also similar to that observed in other species, including tomato, squash and soybean (Minoia et al., 2010; Vicente-Dólera et al., 2014; Tsuda et al., 2015). These data confirm the efficiency of the mutagenesis and demonstrate that the generated mutant library has a high density of mutations, where potential mutant phenotypes can be certainly found by a forward genetic approach.

The frequency of mutant phenotypes in adult plants (28.17%) further encourages the use of our library in forward genetic analyses. Many of them were variations in vegetative

developmental traits such as leave shape and color, and growth habit, but many of the mutant families were altered in male and female flower development (11.97%), including sex determination and sex expression mechanisms (**Table 2**). The high-throughput screening for ethylene insensitivity by using the triple response validate the use of this mutant library in forward genetic approach. A high throughput phenomic screening with more than 35,000 plants resulted in four Ein mutants that showed a reduced response to ethylene in root length and hypocotyl length and thickness. This hormone controls flower and fruit development (Martínez et al., 2013), sexual expression and sex determination (Manzano et al., 2010, 2011, 2013), and chilling injury in refrigerated zucchini fruit (Megías et al., 2016). Therefore, these mutants could be of value not only for investigating the function of ethylene receptor and signaling genes in *C. pepo*, but also for breeding a number of agronomic traits in this and other *Cucurbita* crop species, providing they can be crossed with *C. pepo*. Moreover, knowledge of gene function in *C. pepo* would permit the biotechnological manipulation of other plant species through transgenesis or genome-editing technologies.

Density, Spectrum, Distribution and Impact of EMS Mutations

The mutation density in the squash mutant platform, estimated by WGS of two independent mutant families, was high if compared with other libraries. In addition to sequence quality and coverage, we also filtered SNVs that likely resulted from sequencing and mapping errors (regions with excessive coverage and with high mutation density, probably caused by wrong mapping of repetitive DNA reads), and spontaneous point mutations (SNVs shared by the two families). Discarding clustered SNVs (those that were separated less than 1000 bp) is also in accordance with a study in *Drosophila*, where it was determined that clustered SNPs, which in this case were less than 500 bp apart from one another, were actually natural polymorphisms that arose from gene conversion events (Blumenstiel et al., 2009). After filtering, a total of 2,563 EMS mutations were identified in both L1 and L2 families. The average density was estimated as 1 mutation/111 kb in 144.5 Mb of scored genome, which is slightly higher respect to that so far available and published for a smaller squash mutant population of 768 M2/M3 families (1/133 kb; Vicente-Dólera et al., 2014). The estimated mutation density was also higher in comparison to that published for other EMS mutant collections in the related cucurbit species melon (1 mutation/573 kb; Dahmani-Mardas et al., 2010) and cucumber (1 mutation/1147 kb; Boualem et al., 2014), although in all these cases, the estimations were based on the TILLING of specific target genes and not on WGS. The higher mutation density observed in squash in comparison with other EMS families in related cucurbit species, could be related with the whole genome duplication event observed in *Cucurbita* species (Montero-Pau et al., 2017), which could help in withstanding the mutagen action, as appears to occur in tetraploid and hexaploid wheat, where observed EMS mutation density reaches 1/40 and 1/24 kb, respectively (Slade et al., 2005). The twofold differences in mutation density observed between the two sequenced families has been also reported in mutant libraries of other species like soybean (Tsuda et al., 2015), tomato (Shirasawa et al., 2016), and *Caenorhabditis elegans* (Sarin et al., 2010), where differences among certain families reached up to tenfold.

The most frequent EMS-induced mutations found in the mutant collection were GC to AT transitions, which are known to result from the alkylation of guanine residues. In other mutant libraries of *Arabidopsis*, maize and pea, the frequency of GC to AT transitions represent more than 99% of the EMS mutations (Greene et al., 2003; Till et al., 2004; Dalmais et al., 2008). Whole genome re-sequencing allowed the identification of 80.3 and 61% of GC > AT transitions in L1 and L2, respectively, but other less frequent transitions (TA > CG) and transversions (TA > AT, GC > CG, TA > GC and GC > TA) were also identified in the EMS-induced fraction. Although one could think that these non-canonical mutations are false positives, their occurrence and proportion is consistent with that found in other studies in *Drosophila* (Cooper et al., 2008; Blumenstiel et al., 2009), tomato (Minoia et al., 2010; Shirasawa et al., 2016), soybean (Cooper et al., 2008), rice (Till et al., 2007) and barley (Caldwell et al., 2004), suggesting that, although caused by unknown mechanisms, they are also produced by EMS.

The spectrum of spontaneous mutations occurring in the genome of *C. pepo* was different to that generated by EMS. In this fraction, identified as SNVs shared by the two families under analysis or homozygous for one of the families, the percentage of transitions GC to AT (28.6%) and TA to CG (24.8%) was very similar, while the percentage of transversions reached 46.4%. Moreover, the transition to transversion ratio of spontaneous mutations (1.15) is very different to that in the EMS-induced mutations (6.9 and 3.3 in L1 and L2, respectively), which also strengthen the filtering process followed for the selection of EMS-induced mutations in our study.

The artificial mutations were evenly dispersed throughout the 20 chromosomes of *C. pepo*, and occur randomly in the genome, independently of their location in euchromatic or heterocromatic regions. However, EMS action does not appear to be randomly occurring but is prone to happen in some specific DNA sequences. We found that the two contiguous bases at both sides of the base substitution GC to AT transitions were biased respect to the genome background, with an excess of pyrimidine at −1 and +1 positions, and an excess of C at the −2 position (**Figure 7**). These data are in accordance to that found in other plant species, including tomato (Shirasawa et al., 2016) and *Arabidopsis* (Greene et al., 2003).

CONCLUSION

A large mutation platform has been developed in *C. pepo*, containing 3,751 independent families with an average mutation frequency of 1/111 kb. The mutations are evenly distributed along the complete genome and consist of more than 75% of GC to AT transitions, which are the canonical mutations generated by EMS. The putative impact of the mutation detected by WGS,

and the phenotypic variation observed in the new squash-mutant platform corroborate its usefulness in high-throughput mutation discovery for both gene function analysis and plant breeding. The high-throughput screening of the whole library by using the triple response to ethylene has yielded a total of 4 Ein mutants that are presently being analyzed, and other screenings for biotic and abiotic stresses are underway with the aim to detect novel alleles and new gene functions.

AUTHOR CONTRIBUTIONS

AG conducted most of the experiments. EA collaborated in phenotyping and figures development. SM, CM, and ZM generated the EMS mutant population. GC and JR collaborated in the phenotyping of M2 adult plants. GP and SB performed the bioinformatic analysis and drafted the sequencing methods. MJ and AG interpreted the mutation detection data and drafted the manuscript. DG revised the manuscript draft. MJ coordinated the study and experiments designing.

FUNDING

This work has been supported by grants AGL2014-54598-C2-1-R and AGL2017-82885-C2-1-R, partly funded by the ERDF (European Regional Development Fund) and by the Spanish Ministry of Science and Innovation, and grant P12-AGR-1423, funded by Junta de Andalucía, Spain.

ACKNOWLEDGMENTS

AG, EA, and GC acknowledge FPI, "Garantía Juvenil" and FPU scholarship programs from MEC, Spain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01049/full#supplementary-material

REFERENCES

Bleecker, A. B., Estelle, M. A., Somerville, C., and Kende, H. (1988). Insensitivity to ethylene conferred by a dominant mutation in *Arabidopsis thaliana*. *Science* 241, 1086–1089. doi: 10.1126/science.241.4869.1086

Blumenstiel, J. P., Noll, A. C., Griffiths, J. A., Perera, A. G., Walton, K. N., Gilliland, W. D., et al. (2009). Identification of EMS-induced mutations in *Drosophila melanogaster* by whole-genome sequencing. *Genetics* 182, 25–32. doi: 10.1534/genetics.109.101998

Boualem, A., Fleurier, S., Troadec, C., Audigier, P., Kumar, A. P. K., Chatterjee, M., et al. (2014). Development of a *Cucumis sativus* TILLING platform for forward and reverse genetics. *PLoS One* 9:e97963. doi: 10.1371/journal.pone.0097963

Caldwell, D. G., McCallum, N., Shaw, P., Muehlbauer, G. J., Marshall, D. F., and Waugh, R. (2004). A structured mutant population for forward and reverse genetics in Barley (*Hordeum vulgare* L.). *Plant J.* 40, 143–150. doi: 10.1111/j.1365-313X.2004.02190.x

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46, 714–721. doi: 10.1038/ng.3007

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Colbert, T., Till, B. J., Tompa, R., Reynolds, S., Steine, M. N., Yeung, A. T., et al. (2001). High-throughput screening for induced point mutations. *Plant Physiol.* 126, 480–484. doi: 10.1104/PP.126.2.480

Cooper, J. L., Till, B. J., Laport, R. G., Darlow, M. C., Kleffner, J. M., Jamai, A., et al. (2008). TILLING to detect induced mutations in soybean. *BMC Plant Biol.* 8:9. doi: 10.1186/1471-2229-8-9

Dahmani-Mardas, F., Troadec, C., Boualem, A., Lévêque, S., Alsadon, A. A., Aldoss, A. A., et al. (2010). Engineering melon plants with improved fruit shelf life using the Tilling approach. *PLoS One* 5:e15776. doi: 10.1371/journal.pone.0015776

Dalmais, M., Schmidt, J., Le Signor, C., Moussy, F., Burstin, J., Savois, V., et al. (2008). UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol.* 9:R43. doi: 10.1186/gb-2008-9-2-r43

FAOSTAT (1961). Available at: http://www.fao.org/faostat/en/#home

Fraenkel, R., Kovalski, I., Troadec, C., Bendahmane, A., and Perl-Treves, R. (2014). Development and evaluation of a cucumber TILLING population. *BMC Res. Notes* 7:846. doi: 10.1186/1756-0500-7-846

González, M., Xu, M., Esteras, C., Roig, C., Monforte, A. J., Troadec, C., et al. (2011). Towards a TILLING platform for functional genomics in Piel de Sapo melons. *BMC Res. Notes* 4:289. doi: 10.1186/1756-0500-4-289

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* 164, 731–740.

Guzmán, P., and Ecker, J. R. (1990). Exploiting the triple response of *Arabidopsis* to identify ethylene-related mutants. *Plant Cell* 2, 513–23. doi: 10.1105/tpc.2.6.513

Manzano, S., Martínez, C., Domínguez, V., Avalos, E., Garrido, D., Gómez, P., et al. (2010). A major gene conferring reduced ethylene sensitivity and maleness in *Cucurbita pepo*. *J. Plant Growth Regul.* 29, 73–80. doi: 10.1007/s00344-009-9116-5

Manzano, S., Martínez, C., Megías, Z., Garrido, D., and Jamilena, M. (2013). Involvement of ethylene biosynthesis and signalling in the transition from male to female flowering in the monoecious *Cucurbita pepo*. *J. Plant Growth Regul.* 32, 789–798. doi: 10.1007/s00344-013-9344-6

Manzano, S., Martínez, C., Megías, Z., Gómez, P., Garrido, D., and Jamilena, M. (2011). The role of ethylene and brassinosteroids in the control of sex expression and flower development in *Cucurbita pepo*. *Plant Growth Regul.* 65, 213–221. doi: 10.1007/s10725-011-9589-7

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* 9, 1185–1188. doi: 10.1038/nmeth.2221

Martínez, C., Manzano, S., Megías, Z., Garrido, D., Picó, B., and Jamilena, M. (2013). Involvement of ethylene biosynthesis and signalling in fruit set and early fruit development in zucchini squash (*Cucurbita pepo* L.). *BMC Plant Biol.* 13:139. doi: 10.1186/1471-2229-13-139

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–303. doi: 10.1101/gr.107524.110

Megías, Z., Martínez, C., Manzano, S., García, A., del Mar Rebolloso-Fuentes, M., Valenzuela, J. L., et al. (2016). Ethylene biosynthesis and signaling elements involved in chilling injury and other postharvest quality traits in the non-climacteric fruit of zucchini (*Cucurbita pepo*). *Postharvest Biol. Technol.* 113, 48–57. doi: 10.1016/j.postharvbio.2015.11.001

Minoia, S., Petrozza, A., D'Onofrio, O., Piron, F., Mosca, G., Sozio, G., et al. (2010). A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC Res. Notes* 3:69. doi: 10.1186/1756-0500-3-69

Montero-Pau, J., Blanca, J., Bombarely, A., Ziarsolo, P., Esteras, C., Martí-Gómez, C., et al. (2017). De novo assembly of the zucchini genome reveals a

whole-genome duplication associated with the origin of the *Cucurbita genus*. *Plant Biotechnol. J.* 16, 1161–1171. doi: 10.1111/pbi.12860

Porch, T. G., Blair, M. W., Lariguet, P., Galeano, C., Pankhurst, C. E., and Broughton, W. J. (2009). Generation of a mutant population for TILLING common bean genotype BAT 93. *J. Am. Soc. Hortic. Sci.* 134, 348–355.

Roman, G., Lubarsky, B., Kieber, J. J., Rothenberg, M., and Ecker, J. R. (1995). Genetic analysis of ethylene signal transduction in *Arabidopsis thaliana*: five novel mutant loci integrated into a stress response pathway. *Genetics* 139, 1393–1409.

Sarin, S., Bertrand, V., Bigelow, H., Boyanov, A., Doitsidou, M., Poole, R. J., et al. (2010). Analysis of multiple ethyl methanesulfonate-mutagenized caenorhabditis elegans strains by whole-genome sequencing. *Genetics* 185, 417–430. doi: 10.1534/genetics.110.116319

Sega, G. A. (1984). A review of the genetic effects of ethyl methanesulfonate. *Mutat. Res. Genet. Toxicol.* 134, 113–142. doi: 10.1016/0165-1110(84)90007-1

Shirasawa, K., Hirakawa, H., Nunome, T., Tabata, S., and Isobe, S. (2016). Genome-wide survey of artificial mutations induced by ethyl methanesulfonate and gamma rays in tomato. *Plant Biotechnol. J.* 14, 51–60. doi: 10.1111/pbi.12348

Slade, A. J., Fuerstenberg, S. I., Loeffler, D., Steine, M. N., and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.* 23, 75–81. doi: 10.1038/nbt1043

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Mol. Plant* 10, 1293–1306. doi: 10.1016/j.molp.2017.09.003

Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., et al. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.* 7:19. doi: 10.1186/1471-2229-7-19

Till, B. J., Reynolds, S. H., Weil, C., Springer, N., Burtner, C., Young, K., et al. (2004). Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol.* 4:12. doi: 10.1186/1471-2229-4-12

Tsuda, M., Kaga, A., Anai, T., Shimizu, T., Sayama, T., Takagi, K., et al. (2015). Construction of a high-density mutant library in soybean and development of a mutant retrieval method using amplicon sequencing. *BMC Genomics* 16:1014. doi: 10.1186/s12864-015-2079-y

Vicente-Dólera, N., Troadec, C., Moya, M., Del Rio-Celestino, M., Pomares-Viciana, T., Bendahmane, A., et al. (2014). First TILLING platform in *Cucurbita pepo*: a new mutant resource for gene function and crop improvement. *PLoS One* 9:e112743. doi: 10.1371/journal.pone.0112743

Wang, N., Wang, Y., Tian, F., King, G. J., Zhang, C., Long, Y., et al. (2008). A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. *New Phytol.* 180, 751–765. doi: 10.1111/j.1469-8137.2008.02619.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Journal of Experimental Botany
doi:10.1093/jxb/erz417 Advance Access Publication 28 September 2019
This paper is available online free of all access charges (see https://academic.oup.com/jxb/pages/openaccess for further details)

RESEARCH PAPER

The ethylene receptors *CpETR1A* and *CpETR2B* cooperate in the control of sex determination in *Cucurbita pepo*

Alicia García[1], Encarnación Aguado[1], Cecilia Martínez[1], Damian Loska[2], Sergi Beltrán[2], Juan Luis Valenzuela[1], Dolores Garrido[3], and Manuel Jamilena[1,*]

[1] Department of Biology and Geology, Research Centers CIAIMBITAL and CeiA3, University of Almería, 04120 Almería, Spain
[2] Centro Nacional de Análisis Genómico (CNAG), 08028 Barcelona, Spain
[3] Department of Plant Physiology, University of Granada, 18071 Granada, Spain.

* Corresponding author: mjamille@ual.es

Received 23 April 2019; Editorial decision 4 September 2019; Accepted 10 September 2019

Editor: Frank Wellmer, Trinity College Dublin, Ireland

Abstract

High-throughput screening of an ethyl methanesulfonate-generated mutant collection of *Cucurbita pepo* using the ethylene triple-response test resulted in the identification of two semi-dominant ethylene-insensitive mutants: *etr1a* and *etr2b*. Both mutations altered sex determination mechanisms, promoting conversion of female into bisexual or hermaphrodite flowers, and monoecy into andromonoecy, thereby delaying the transition to female flowering and reducing the number of pistillate flowers per plant. The mutations also altered the growth rate and maturity of petals and carpels in pistillate flowers, lengthening the time required for flowers to reach anthesis, as well as stimulating the growth rate of ovaries and the parthenocarpic development of fruits. Whole-genome sequencing allowed identification of the causal mutation of the phenotypes as two missense mutations in the coding region of *CpETR1A* and *CpETR2B*, each one corresponding to one of the duplicates of ethylene receptor genes highly homologous to Arabidopsis *ETR1* and *ETR2*. The phenotypes of homozygous and heterozygous single- and double-mutant plants indicated that the two ethylene receptors cooperate in the control of the ethylene response. The level of ethylene insensitivity, which was determined by the strength of each mutant allele and the dose of wild-type and mutant *etr1a* and *etr2b* alleles, correlated with the degree of phenotypic changes in the mutants.

Keywords: *Cucurbita pepo*, ethylene, ethyl methanesulfonate, mutants, sex determination, fruit set.

Introduction

Monoecious and dioecious plant species produce unisexual flowers (male or female) either in the same plant (monoecy) or in separate plants (dioecy). They are believed to be derived from a hermaphrodite ancestor, by different mechanisms, which result in the suppression of either stamen or carpel primordia development during the formation of female or male flowers, respectively (Dellaporta and Calderon-Urrea, 1993; Pannell, 2017). The genetic control mechanisms underlying sex determination in plants are diverse, ranging from heteromorphic sex chromosomes, as occurs in the dioecious species *Silene latifolia* and *Rumex acetosa*, to a number of non-linked genes, as occurs in the monoecious species of the family Cucurbitaceae (Jamilena *et al.*, 2008; Pannell, 2017)

Cultivars of Cucurbitaceae species, including *Cucumis sativus* (cucumber), *Cucumis melo* (melon), *Citrullus lanatus* (watermelon), and the species of the genus *Cucurbita* (pumpkins and

many squashes), are mainly monoecious, although certain cultivars of *C. sativus*, *C. melo*, and *C. lanatus* are andromonoecious, producing male and hermaphrodite flowers on the same plant (Malepszy and Niemirowicz-Szczytt, 1991; Perl-Treves, 2004; Boualem *et al.*, 2009; Manzano *et al.*, 2016). Both monoecious and andromonoecious plants go through two flowering phases of development: an initial phase in which the plant produces only male flowers, and a second phase in which the plant alternates the production of pistillate and male flowers (Perl-Treves 2004; Manzano *et al.*, 2013, 2014; Zhang *et al*, 2017). The transition to pistillate flowering, and the number of pistillate flowers per plant, vary within the different cultivars of each species. In *Cucumis*, natural genetic variation includes other sexual phenotypes, such as gynoecious (a plant producing only female flowers) and androecious (a plant producing only male flowers). Neither the andromonoecious phenotype nor the gynoecious and androecious phenotypes have been observed in the genus *Cucurbita*, although some cultivars show a partially andromonoecious phenotype characterized by the occurrence of male and bisexual flowers, that is, pistillate flowers with partially developed stamens and no pollen (Martínez *et al.*, 2014).

Sex determination mechanisms in cucurbit species are controlled by ethylene. Under treatments that reduce ethylene biosynthesis or perception, the monoecious plants are converted into partially or completely andromonoecious ones, demonstrating that ethylene participates in the sex identity of female flowers, and that an ethylene threshold is required to arrest stamen development in the female flower (Manzano *et al.*, 2011; Zhang *et al.*, 2017). Besides the control of individual floral buds, ethylene also participates in the control of sex expression within the plant. A reduction in ethylene biosynthesis or perception delays the transition to pistillate flowering and reduces the number of pistillate flowers per plant in *C. sativus, C. melo*, and *C. pepo*, but has the opposite effect in *C. lanatus* (Manzano *et al.*, 2011, 2014). By contrast, treatment with ethylene or ethylene-releasing agents induces the production of male flowers in *C. lanatus* but promotes the production of pistillate flowers in the other species (Rudich *et al.*, 1969; Byers *et al.*, 1972; Den Nijs and Visser, 1980).

The genes and mutations responsible for cucurbit sex phenotypes are currently being sought. The arrest of stamen development in the female flowers of the different species requires the functioning of the ethylene biosynthesis orthologs *CmACS7*, *CsACS2*, *CpACS27A*, and *CitACS4*, which are specifically expressed in the female flowers of *C. melo*, *C. sativus*, *C. pepo*, and *C. lanatus*, respectively. Loss-of-function mutations in these genes led to andromonoecy in *C. melo*, *C. sativus*, and *C. lanatus* (Boualem *et al.*, 2008, 2009; Ji *et al.*, 2016; Manzano *et al.*, 2016), but to only partial andromonoecy in *C. pepo* (Martínez *et al.*, 2014). The androecious and gynoecious phenotypes also resulted from two independent mutations: in *C. melo*, androecy resulted from a mutation in the *CmACS11* gene (Boualem *et al.*, 2015), while gynoecy was produced by a mutation in the *CmWIP1* gene (Martin *et al.*, 2009). *CmACS11* represses the expression of *CmWIP1* to permit the coexistence of male and female flowers in monoecious species (Boualem *et al.*, 2015).

Having performed extensive screening of an ethyl methanesulfonate (EMS)-generated mutant collection of *C. pepo* in the search for ethylene-insensitive mutants (García *et al.*, 2018) in order to gain further insights into the genetic network regulating sex determination in cucurbits, in this paper we present a molecular and functional characterization of two semi-dominant mutations that affect two ethylene receptors of *C. pepo*: *CpETR1A* and *CpETR2B*. These mutations confer ethylene insensitivity on the plant, resulting in the conversion of female flowers to bisexual or hermaphrodite ones; that is, monoecy to andromonoecy. The mutations also alter the development and maturity of different floral organs in pistillate flowers, including ovaries and fruit.

Materials and methods

Plant material

The ethylene-insensitive mutants analyzed in this study were selected from a high-throughput screening of a *C. pepo* mutant collection by the triple-response assay (García *et al.*, 2018), consisting of shortening and thickening of hypocotyls and roots in seedlings germinated in the dark with an external input of ethylene (Bleecker *et al.*, 1988). The *etr1a* and *etr2b* mutants analyzed in this paper correspond to the *ein2* and *ein3* mutants isolated by García *et al.* (2018).

Before phenotyping, *etr* mutant plants from each family were crossed for two generations with the background genotype MUC16, and the resulting BC_2 generation was selfed to obtain the BC_2S_1 generation. Given that homozygous *etr1a* and *etr2b* mutants were female sterile, they were always derived from selfed progenies of BC heterozygous plants. Sterility also prevented us from obtaining double homozygous mutants. The heterozygous double mutants (wt/*etr1a* wt/*etr2b*) were obtained by crossing heterozygous wt/*etr1a* as female and homozygous *etr2b*/*etr2b* as male, and genotyping the offspring for the causal mutations. The double heterozygous plants were also female sterile, preventing us from obtaining the double homozygous *etr1a*/*etr1a etr2b*/*etr2b*.

Phenotyping for sex expression and sex determination traits

First, BC_1S_1 or BC_2S_1 plants from *etr1a* and *etr2b* mutant families were classified according to their level of triple response to ethylene in wild-type (WT), intermediate (wt/*etr*), and ethylene-insensitive mutants (*etr*/*etr*), and then transplanted to a greenhouse and grown to maturity under local greenhouse conditions without climate control, and under standard crop management, in Almería, Spain.

For the ethylene response assay, seeds were germinated for 2 days in the absence of ethylene and then placed in a growth chamber containing 50 ppm ethylene in darkness for 5 days. The identified mutants showed more elongated hypocotyls and roots than WT, resembling seedlings grown in air. The ethylene sensitivity of each mutant genotype was estimated by using three replicates with at least 20 seedlings of the same genotype. Ethylene sensitivity was assessed as the percentage of reduction in hypocotyl length relative to air-grown seedlings: $[(H_0–H_1/H_0)\times100]$, where H_0 corresponds to the hypocotyl length of air-grown seedlings, and H_1 to the hypocotyl length of ethylene-treated seedlings. The final ethylene sensitivity was relativized to the WT seedling response, considering that the WT has 100% ethylene sensitivity, and ethylene insensitivity was calculated as (100–ethylene sensitivity).

The sex phenotype of each plant was determined according to the sex of the flowers in the first 40 nodes of each plant. A minimum of 30 WT, 30 wt/*etr*, and 30 *etr* plants were phenotyped for each mutant family. The sex expression of each genotype was assessed by determining the node at which plants transitioned to pistillate flowering, and the number of male or pistillate flower nodes. The sex phenotype of each individual pistillate flower was assessed by the so-called andromonoecy index (AI) (Martínez *et al.*, 2014; Manzano *et al.*, 2016). Pistillate flowers were separated into three phenotypic classes that were given a score from 1 to 3 according to the degree of their stamen development: female (AI=1), showing no stamen development; bisexual (AI=2), showing partial development of stamens and

no pollen; and hermaphrodite (AI=3), showing complete development of stamens and pollen. The average AI of each plant and genotype was then assessed from the resulting AI score of at least five individual pistillate flowers from each plant, using a minimum of 30 plants for each genotype.

The growth rates of ovaries and petals in both pistillate and male flowers in each of the WT and mutant plants were assessed by measuring the length of ovaries and petals every 3 days in at least 12 flowers of each genotype, starting with flowers ~2 mm in length. The anthesis time was estimated as the number of days taken for a 2 mm pistillate or male floral bud to reach anthesis. The effect of the *etr1a* and *etr2b* mutations on the vegetative vigor of each plant was assessed by determining the average plant height, the total number of nodes, and the average internode length in the main shoot of WT and mutant plants at 60 days after transplantation.

Identification of etr1a and etr2b mutations by whole-genome sequencing analysis

To identify the causal mutations of the *etr1a* and *etr2b* phenotypes, WT and mutant plants, which were both derived from BC_2S_1 segregating populations, were subjected to whole-genome sequencing (WGS). In total, 120 BC_2S_1 seedlings from each mutant family were subjected to the ethylene triple-response assay, and the plants exhibiting the ethylene-insensitive mutant phenotype were separated from the WT. The phenotype of those seedlings was verified in the adult plants, since WT plants were monoecious while homozygous mutant plants were andromonoecious or partially andromonoecious.

The DNA of each plant was isolated by using the Gene JET Genomic DNA Purification Kit (Thermo Fisher), and the DNA from 20 plants showing the same phenotype in each mutant family were pooled into different bulks. Two DNA bulks were generated for each mutant family: one WT bulk and one mutant bulk. The four DNA bulks were used to prepare paired-end multiplex libraries by using the KAPA Library Preparation kit (Kapa Biosystems) and sequenced in 126 base-read pairs on an Illumina HiSeq2000 instrument, following the manufacturer's protocol. Base calling and quality control were done using the Illumina RTA sequence analysis pipeline, according to the manufacturer's instructions. The average fold-effective coverage of all the samples was between 15.5 and 17.2, and the percentage of genomic bases with a fold coverage higher than 5 was between 78.4 and 83.3% (see Supplementary Table S1 at *JXB* online). Raw reads were mapped on to the *C. pepo* genome v3.2 by using the GEM mapper (Marco-Sola *et al.*, 2012) in order to generate four bam files. Indels were realigned with GATK, and duplicates were marked with Picard v1.110. In each sample we counted all the 'good-quality bases', selected according to the following filter parameters: base quality ≥17, mapping quality ≥20, read depth ≥8. Variants were called using GATK's HaplotypeCaller in gVCF mode (McKenna *et al.*, 2010), and then the GenotypeGVCF tool was used to perform joint genotyping on all the samples together. Only the single nucleotide polymorphisms (SNPs) were selected by using GATK's SelectVariants tool. Sites with multiple nucleotide polymorphisms were filtered out.

With regard to the segregation analysis of each mutant phenotype, the genotype of WT plants should be considered as 0/0, as the reference genome. The mutant bulks were completed by using DNA from the most ethylene-insensitive plants. However, given the semi-dominant nature of the *etr1a* and *etr2b* mutations, we expected some plants to be heterozygous for the mutations (0/1), although most of them were expected to be homozygous (1/1), with a mutant allelic frequency (AF) close to 1. Therefore, we filtered our data according to the following parameters: genotype quality ≥20, read depth ≥4 (this specific base was covered with at least four reads), AF ≥0.8 in the two mutant bulks, and AF ≤0.2 in the WT bulks.

Validation of the identified mutations by high-throughput genotyping of individual segregating plants

To confirm that the identified mutations were the causal mutations of the *etr1a* and *etr2b* phenotypes, more than 200 individual BC_2S_1 segregating populations were genotyped by using real-time PCR with TaqMan probes. The multiplex PCRs were done using the Bioline SensiFAST™ Probe No-ROX Kit, a set of forward and reverse primers amplifying the polymorphic sequence, and two allele-specific probes descriptive of the SNP of interest (C–T). The WT probe was labeled with FAM dye, while the mutant probe was labeled with HEX reporter dye. The BHQ1 quencher molecule was used in both probes (Supplementary Table S2). After genotyping, plants were phenotyped for ethylene triple response in etiolated seedlings and for sex expression in adult plants, enabling us to see whether the mutant alleles co-segregated with the *etr1a* or the *etr2b* mutant phenotype.

Assessment of relative gene expression by quantitative RT–PCR

Gene expression analysis was performed by using quantitative reverse transcription (qRT)–PCR on three biological replicates, each one resulting from an independent extraction of total RNA from samples that were pooled from at least three plants with the same genotype.

RNA was isolated according to the protocol of the GeneJET Plant RNA Purification Kit (Thermo Fisher). cDNA was then synthesized by using the RevertAid RT Reverse Transcription Kit (Thermo Fisher). The expression levels of genes were evaluated through qRT–PCR by using the Rotor-Gene Q thermocycler (Qiagen) and SYBR® Green Master Mix (BioRad). Supplementary Table S3 shows the primers used for qRT–PCR reactions for each analyzed gene.

Bioinformatics and statistical analyses

Alignments were performed using the BLAST alignment tools at NCBI (http://www.blast.ncbi.nlm.nih.gov/) and Clustal Omega at EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/).

The phylogenetic relations of ETR1, ETR2, and ERS1-like ethylene receptors were studied using MEGA7 software (Kumar *et al.*, 2016), which allowed the alignment of proteins and the construction of phylogenetic trees using the Maximum Likelihood method based on the Poisson correction model (Zuckerkandl and Pauling, 1965), with 2000 bootstrap replicates. The protein sequences (Supplementary Table S4) were obtained using The Arabidopsis Information Resource (https://www.arabidopsis.org/) and the Cucurbit Genomics Database (http://cucurbitgenomics.org/).

Multiple data comparisons were obtained by analysis of variance with the significance level $P<0.05$, and each two averages were compared using Fisher's least significant difference method.

Results

etr1a *and* etr2b *are two semi-dominant ethylene-insensitive mutations affecting sex determination*

etr1a and *etr2b* are two independent, ethylene-insensitive mutant families that were found after screening a mutant collection of *C. pepo* for ethylene triple response (García *et al.*, 2018). To ensure accurate phenotyping, mutant plants were backcrossed with the background genotype MUC16 for two generations, and then selfed. The resulting BC_1S_1 or BC_2S_1 generations segregated for three ethylene-response phenotypes in etiolated seedlings: ethylene-sensitive plants (WT), intermediate plants (wt/*etr*), and ethylene-insensitive plants (*etr*) (Fig. 1). The segregation ratio of the three ethylene triple-response phenotypes in BC_1S_1 and BC_2S_1 generations indicated that the two ethylene-insensitive mutations were semi-dominant, and that the intermediate phenotype corresponded to heterozygous plants (wt/*etr1a* or wt/*etr2b*) (Supplementary Table S5) (García *et al.*, 2018).

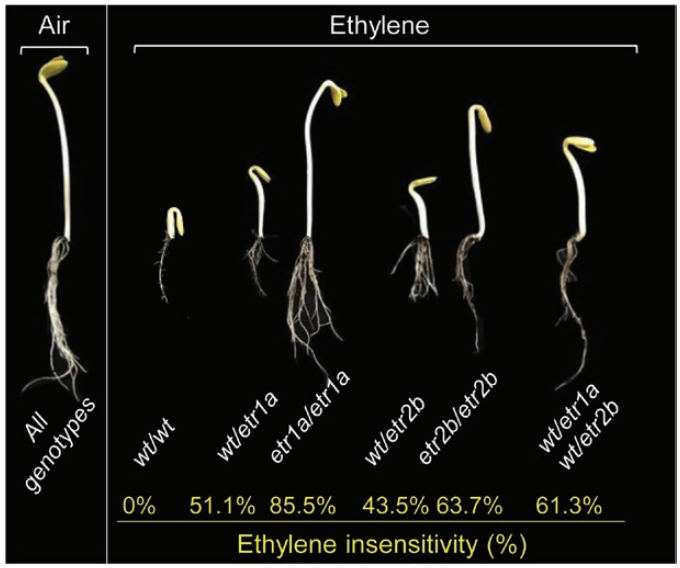


Fig. 1. Ethylene triple-response phenotypes of WT and heterozygous and homozygous single and double mutants for *etr1a* and *etr2b*. When grown in air, both the WT and the mutants showed the same growth. When exposed to ethylene, the WT responded with drastic reductions in the length of the hypocotyl and roots, while the mutants had a minor but differential response. The average level of ethylene sensitivity (ES) was assessed as the percentage of reduction in hypocotyl length relative to air-grown seedlings, and assuming that WT is 100% ethylene sensitive. The percentage ethylene insensitivity was then calculated as (100–ES). Assessments were performed in three replicates with at least 20 seedlings for each genotype.

WT plants responded to ethylene with reduced length of the roots and the hypocotyl, and also increased hypocotyl thickness, in comparison to those grown in air. In contrast, the length of the hypocotyl and roots of ethylene-treated homozygous *etr1a* and heterozygous *etr2b* seedlings was reduced to a lesser extent in relation to those grown in air (Fig. 1); this indicated that the level of insensitivity of the two mutants was not total, as they possessed a partially ethylene-insensitive phenotype. Assuming that WT plants are completely ethylene sensitive (0% ethylene insensitivity), we estimated the percentage of ethylene insensitivity in homozygous and heterozygous seedlings of the two mutants. The homozygous *etr1a* seedlings showed a more severe ethylene-insensitive phenotype than the homozygous *etr2b* seedlings (85.5% ethylene insensitivity for *etr1a*/*etr1a* versus 63.7% in *etr2b*/*etr2b*). Heterozygous mutants displayed an intermediate percentage of ethylene insensitivity (51.1% in wt/*etr1a* and 43.5% in wt/*etr2b*) (Fig. 1).

Given that ethylene is the main regulator of sex determination in cucurbits, we assessed whether the *etr1a* and *etr2b* mutations altered the sexual phenotype of the plants. Staminate and pistillate flowers in the first 40 nodes of WT, heterozygous, and homozygous mutant plants were assessed (Supplementary Fig. S1). All plants showed two sexual phases of development. In the first phase, plants produce only male flowers. The second phase starts after the pistillate flowering transition and is characterized by the production of male and pistillate flowers alternately (Fig. 2A). The duration of these two phases of sexual development was affected by the *etr* mutations. Homozygous *etr1a* and *etr2b*

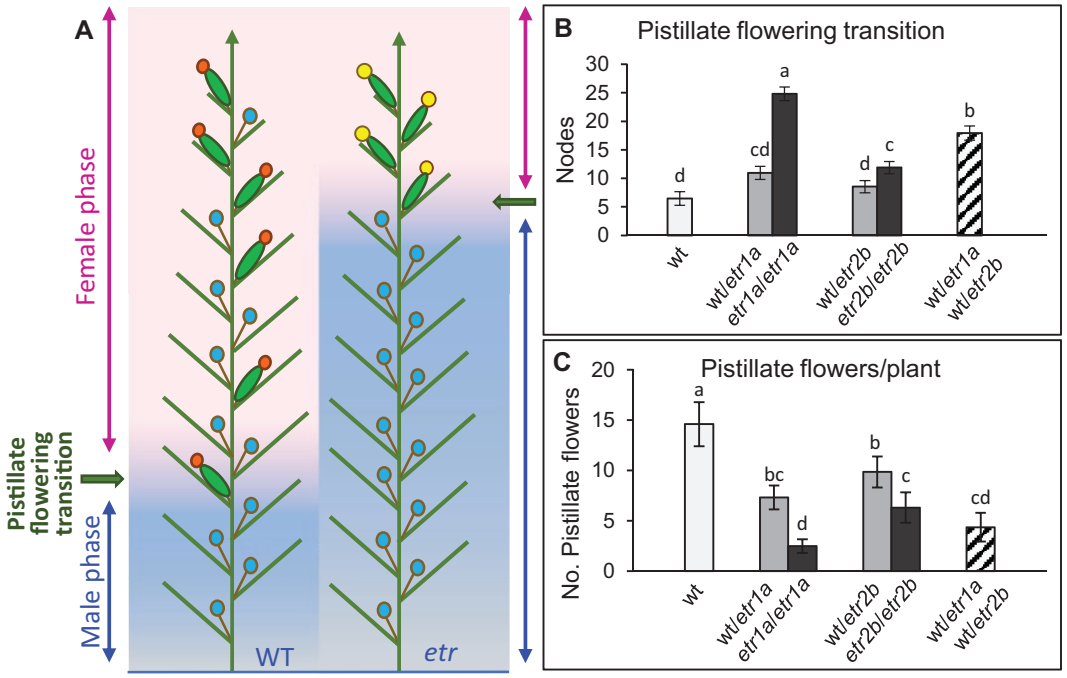


Fig. 2. Effect of *etr1a* and *etr2b* on sex expression. (A) Schematic representation of the distribution of male and female flowers in WT and *etr1a* or *etr2b* (*etr*) mutant plants. The male and female phases of development, and the node at which plants start to produce pistillate flowers (pistillate flowering transition), are indicated. Blue = male flower; red = female flower; yellow = bisexual or hermaphrodite flower. (B) Comparison of pistillate flowering transition in WT plants and single and double mutants. (C) Comparison of the number of pistillate flowers per plant in WT and single and double mutant plants. Error bars represent SE. Different letters in (B) and (C) indicate statistically significant differences ($P<0.05$) between samples.

mutants showed a significant delay in pistillate flowering transition, as well as a reduction in the number of pistillate flowers per plant (Fig. 2B, C). Heterozygous *etr1a* and *etr2b* mutants showed an intermediate phenotype for both traits (Fig. 2B, C), confirming the semi-dominant nature of the two mutations.

The most evident phenotypic alteration in the *etr1a* and *etr2b* mutants was the conversion of monoecy into partial or complete andromonoecy. WT plants produced exclusively female flowers after the pistillate flowering transition, indicating a complete arrest of stamen development in pistillate flowers (Fig. 3A). In the homozygous *etr1a* mutants, almost all female flowers were transformed into hermaphrodite flowers, showing fully developed stamens with fertile pollen (Fig. 3B). In *etr2b* mutants, a high percentage of female flowers were transformed into bisexual (pistillate flowers with immature stamens) or hermaphrodite flowers; however, a small number of female flowers remained (Fig. 3B). In heterozygous plants bearing both mutations, only a very small proportion of the female flowers were converted into bisexual flowers (Fig. 3B), indicating that with regard to the sex of the flower, *etr1a* and *etr2b* mutations are recessive, and therefore not semi-dominant as shown for ethylene triple response and sex expression. Given that the degree of stamen development in pistillate flowers was variable, plants were classified according to the AI, ranging from AI=1 (monoecious) to AI=3 (andromonoecious) (Fig. 3A). All ethylene-sensitive WT plants showed an average AI of 1. Homozygous plants containing either *etr1a* or *etr2b* mutations displayed an average AI of 2.8 and 1.8, respectively, indicating that the *etr1a* mutation rendered plants almost completely andromonoecious, while the *etr2b* mutation rendered them partially andromonoecious. The average AI for heterozygous plants was 1.1 for both *etr1a* and *etr2b* types, very similar to that of WT plants (Fig. 3C).

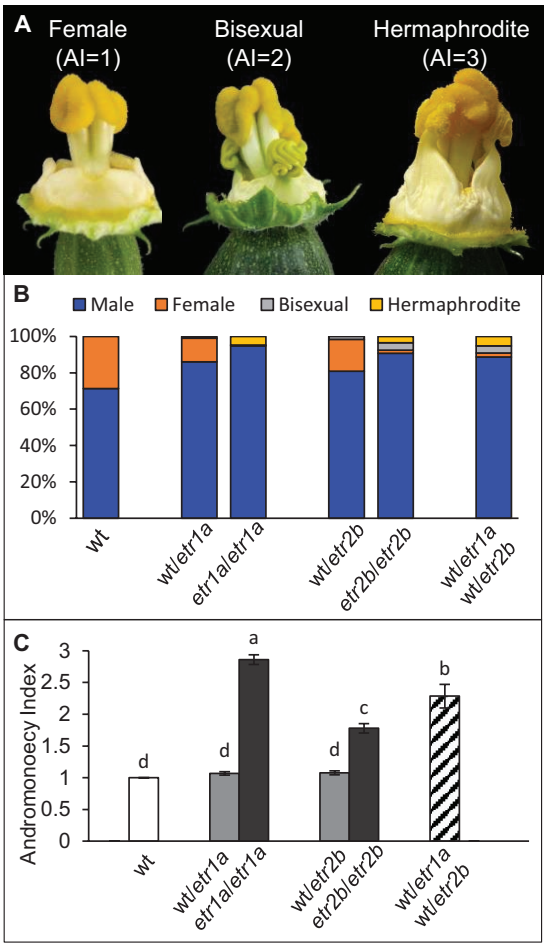


Fig. 3. Effect of *etr1a* and *etr2b* mutations on sex determination in *C. pepo*. (A) Phenotypes of pistillate flowers in WT and mutant plants: female [andromonoecy index (AI)=1], bisexual (AI=2), and hermaphrodite (AI=3). Female flowers develop no stamen, while bisexual and hermaphrodite flowers develop immature stamens and mature stamens with pollen, respectively. (B) Percentage of male, female, bisexual, and hermaphrodite flowers in WT plants and *etr1a* and *etr2b* single and double mutants. (C) AI of WT plants and *etr1a* and *etr2b* single and double mutants. The average AI of each genotype was assessed in at least five individual pistillate flowers from each plant, with each genotype containing a minimum of 30 plants. AI varies between 1 (complete monoecy) and 3 (complete andromonoecy). Error bars represent SE. Different letters indicate statistically significant differences ($P<0.05$) between samples.

etr1a *and* etr2b *alter petal and ovary/fruit development and affect plant vigor*

Table 1 and Fig. 4 show the effects of the *etr1a* and *etr2b* mutations on petal and ovary/fruit development. In the hermaphrodite flowers of *etr1a*, the petal growth rate was reduced and resembled petal development in male flowers. Petal maturity and subsequent anthesis of the flower were delayed (Fig. 4B). Anthesis time, which is the period of time taken for a 2 mm floral bud to reach anthesis and to open, was longer in male WT flowers (average 22 days) than in female WT flowers (average 14 days) (Table 1). Hermaphrodite *etr1a* flowers took an average of 22 days to reach anthesis (range 15–40 days). Under the greenhouse conditions used, several hermaphrodite flowers did not reach anthesis; the petals remained green and closed for more than 40 days (Fig. 4B). The development to maturity of petals in bisexual flowers of *etr2b* was also delayed in comparison with WT female flowers. However, the delay was less pronounced than in *etr1a* (Fig. 4B, Table 1). No alterations in petal development or anthesis time were observed in *etr1a* and *etr2b* male flowers, and no change was found in petal development or anthesis time of female flowers of *etr1a* or *etr2b* heterozygous plants (Fig. 4B; Table 1).

Significant differences in ovary size were detected between WT and *etr1a* and *etr2b* pistillate flowers (Fig. 4). While WT ovaries reached ~8 cm at anthesis and then aborted, the flowers of *etr1a* and *etr2b* reached 16–40 cm (or even more) at anthesis (Fig. 4C). These can be considered as parthenocarpic fruits because the

Table 1. *Anthesis time of pistillate and male flowers in* etr1a *and* etr2b *mutant families*

Family	Flower	Anthesis time (days)		
		wt/wt	**wt/*etr***	***etr/etr***
etr1a	Pistillate	14.3±1.2 [b]	14.8±1 [b]	22.7±1.4 [a]
	Male	22.4±0.7 [a]	23.1±1.8 [a]	22.7±1.8 [a]
etr2b	Pistillate	13.8±1.3 [b]	14.3±1.1 [b]	20±4.2 [a]
	Male	22.4±0.5 [a]	21.6±1.1 [a]	22±1.7 [a]

Different letters within the same row indicate significant differences between means ($P<0.05$).

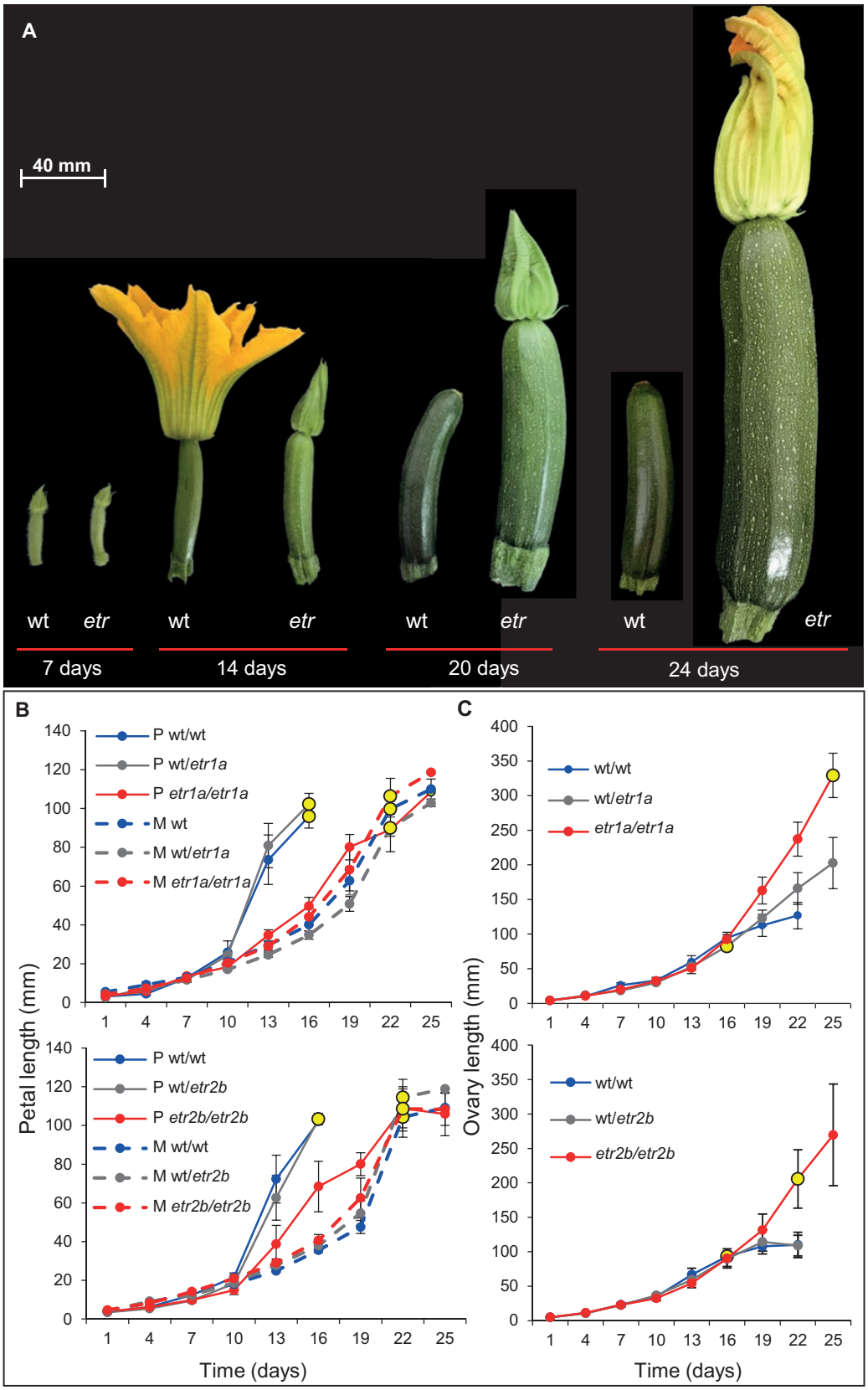


Fig. 4. Effect of *etr1a* and *etr2b* mutations on the growth rate of petals and ovaries of male and pistillate flowers. (A) Images of pistillate flowers in WT and either *etr1a* or *etr2b* mutants (*etr*) at 7, 14, 20, or 24 days after the flower's ovary reached 4 mm in length. The WT pistillate flower reached anthesis at ~15 days, and in the absence of pollination, the floral organs abscised and the fruit aborted 2–3 days after anthesis. In the two *etr* mutants, the anthesis time was markedly delayed, but the fruit grew normally in the absence of pollination (parthenocarpic fruits). (B) Comparison of the growth rate of WT and mutant petals. Flowers were labeled when their ovaries were 4 mm long, and then measured every 3 days for 25 days. P = pistillate flowers, M = male flowers. Yellow circles indicate the time at which more than 80% of the flowers reached anthesis. (C) Comparison of the growth rate of WT and mutant ovaries/fruits over a period of 25 days. Error bars represent SE.

flowers remained closed, rendering them unavailable for pollination. During the first 16 days, the growth rates of WT and mutant ovaries were similar. After 16 days, WT ovaries aborted, while those of homozygous *etr1a* and *etr2b* flowers maintained their growth up to anthesis. Since the anthesis time of *etr1a* and *etr2b* pistillate flowers was achieved much later than that of WT plants, the size of the mutant ovary at anthesis was much larger (Fig. 4C). The ovary growth rate of heterozygous wt/*etr1a* flowers was intermediate between that of wt/wt and homozygous *etr1a*/*etr1a* (Fig. 4C), while heterozygous wt/*etr2b* ovaries displayed the same growth rate as WT plants. Pollination was attempted in homozygous mutant flowers that reached anthesis, but none of the fruits in the homozygous *etr1a* and *etr2b* mutant plants were able to set seeds under the given conditions. Since the pollen of these mutants is fertile, these results indicate female sterility associated with the *etr1a* and *etr2b* mutations. As no pure seed could be obtained from *etr1a* or *etr2b* homozygous mutants, they were maintained by selfing heterozygous mutant plants.

Vegetative development was enhanced in mutant adult *etr1a*/*etr1a* and *etr2b*/*etr2b* plants. Table 2 shows the differences in plant height, number of nodes, and internode length of plants grown under the same conditions. The two mutations were associated with increased plant growth rate and height compared with WT plants. These differences in height and growth rate were mainly due to an increase in the internode length, which was approximately twice as high in mutants as in WT plants. The number of nodes developed by WT and mutant plants was, however, very similar (Table 2).

Phenotype of etr1a *and* etr2b *double mutants*

The interaction between the *etr1a* and *etr2b* mutations was studied in double-heterozygous plants for both mutations. As mutant bisexual and hermaphrodite flowers were sterile, heterozygous wt/*etr1a* plants were pollinated with pollen from homozygous *etr2b*/*etr2b* plants. The progeny segregated as 1:1 for single-heterozygous wt/wt wt/*etr2b* and double-heterozygous wt/*etr1a* wt/*etr2b* plants. However, double-heterozygous plants were also female sterile, which made it impossible to generate double-homozygous plants bearing the two mutations. The reciprocal cross produced the same results.

Double-heterozygous mutants showed a more severe ethylene-insensitive phenotype (61.3% insensitivity) than single-heterozygous mutants (51.1% insensitivity for wt/*etr1a* and 43.5% for wt/*etr2b*), and their phenotype resembled the triple response of *etr2b*/*etr2b* (Fig. 1). The maleness effect of the *etr1a* and *etr2b* mutations was also enhanced in the double-heterozygous wt/*etr1a* wt/*etr2b* plants compared with the phenotype of the single-heterozygous mutants. The female flowering transition was delayed up to an average of 17.9 nodes, and pistillate flower production was also significantly reduced to about 4.3 flowers resulting from the first 40 nodes of the plant, again resembling the phenotype exhibited by the homozygous single mutants (Fig. 2B, C).

In the double-heterozygous wt/*etr1a* wt/*etr2b* plants, most female flowers were converted into bisexual or hermaphrodite flowers (average AI=2.3), contrasting with single-heterozygous plants, which produced nearly 100% female flowers (average AI=1.1) (Fig. 3B, C). Thus, the combination of the two *etr* mutations, albeit in heterozygous conditions, had a similar effect on the flower sexual phenotype as the homozygous single mutations (Fig. 3B, C). These data therefore indicate that the two mutations have an additive effect on the sex phenotype of pistillate flowers, and that the effect is dependent upon the number of mutant alleles for the two loci.

Table 2. *Effects of the* etr1a *and* etr2b *mutations on plant vegetative development*

	Plant height (cm)	Node number	Internode length (mm)
WT	85.43 [b]	47.2 [b]	13.5 [b]
etr1a/etr1a	112.5 [a]	50.7 [a]	24.1 [a]
WT	93.9 [b]	47.1 [a]	14.8 [b]
etr2b/etr2b	113.3 [a]	48.5 [a]	28.8 [a]

Different letters indicate significant differences between wild-type (WT) and homozygous mutants ($P<0.05$).

Identification of the etr1a *and* etr2b *mutations*

The causal mutations of the *etr1a* and *etr2b* phenotypes were identified by WGS. One DNA WT bulk and one DNA mutant bulk were made for each mutant family, each bulk consisting of a pool of DNA from 20 plants from the same BC_2S_1 segregating population. With regard to the mutant bulk, only the ethylene-insensitive plants that showed the most strongly andromonoecious phenotype were selected, in order to assure that they were homozygous for the mutations. The four DNA bulks were subjected to WGS and the results were mapped against the *C. pepo* reference genome version 3.2. The identified SNPs were filtered by using different criteria. Common variants between samples of the two families were discarded, as they were considered likely to correspond to spontaneous nucleotide polymorphisms in the MUC16 genetic background (García *et al.*, 2018). SNPs corresponding to canonical EMS mutations (C>T and G>A) were selected, and filtered for their quality and depth, as well as for their mutant allele frequency (AF), in WT and mutant DNA samples. We expected the WT bulks to present the genotype 0/0 (AF=0) and the mutant bulks to present the genotype 1/1 (AF=1).

Assuming minor contamination of the bulks with some heterozygous plants, we discarded SNPs with AF<0.8 in the mutant samples and AF>0.2 in the WT samples. After filtering, two EMS candidate mutations were selected for *etr1a* and one candidate mutation was selected for *etr2b* (Supplementary Table S1). The sequences surrounding the candidate mutations (±500 bp) were then used in BLAST searches against the DNA and protein databases at NCBI, which detected an EMS canonical C>T transition in both *etr1a* and *etr2b* families, located in the coding region of the ethylene receptor genes *CpETR1A* and *CpETR2B*, respectively (Table 3). To verify that these were the causal mutations of the *etr1a* and *etr2b* phenotypes, more than 200 plants segregating for either *etr1a* or *etr2b* were genotyped for the WT and mutant alleles of *CpETR1A* and *CpETR2B*. The results demonstrated a perfect co-segregation of the *etr1a* and *etr2b* sex phenotypes with

Table 3. *Genotype, depth (DP), and allele frequency (AF) of causal mutations of the* etr1a *and* etr2b *phenotypes*

	DNA bulk	Causal mutations					
		LG07: 6 891 436 (C>T)			LG03:11 824 350 (C>T)		
		Genotype	DP	AF	Genotype	DP	AF
etr1a	WT	0/0	20	0.15	0/0	19	0
	Mutant	1/1	20	1	0/0	9	0
etr2b	WT	0/0	16	0	0/0	10	0
	Mutant	0/0	8	0	1/1	14	1

Table 4. *Co-segregation analysis of the* CpETR1A *and* CpETR2B *mutations with the ethylene-insensitive phenotypes in* BC_2S_1 *populations segregating for* etr1a*,* etr2b*, and double mutants*

Segregating population	*CpETR1A* or *CpETR2B* genotypes	Triple response to ethylene			Sexual phenotype		Total
		Sensitive	Intermediate	Insensitive	Monoecious	Andromonoecious	
etr1a	0/0	80	–	–	80		226
	0/1	–	61		61		
	1/1	–		85		85	
etr2b	0/0	85	–	–	85		234
	0/1	–	62		62		
	1/1	–		87		87	
Double mutants	0/0; 0/1	–	22		22		47
	0/1; 0/1	–		25		25	

mutations in *CpETR1A* and *CpETR2B*, respectively (Table 4). The other candidate mutation for *etr1a* segregated independently of the mutant phenotype.

With regard to the ethylene triple response, the plants that were homozygous for the WT alleles were sensitive to ethylene, and those that were homozygous for the mutant alleles of *CpETR1A* or *CpETR2B* were ethylene insensitive; heterozygous plants showed intermediate triple-response phenotypes (Table 4). Moreover, plants that were homozygous for any of the identified mutations were all andromonoecious, while those that were either homozygous for the WT allele or heterozygous were all monoecious (Table 4). This finding demonstrates that mutations identified in the two ethylene receptor genes *CpETR1A* and *CpETR2B* co-segregated with the ethylene-insensitive phenotype and andromonoecy of the *etr1a* and *etr2b* mutants.

Gene structure of CpETR1 *and* CpETR2

The *de novo* assembly of the *C. pepo* genome, published recently, revealed a whole-genome duplication, which occurred just before the speciation event that created the genus *Cucurbita* (Sun *et al.*, 2017; Montero-Pau *et al.*, 2018). Accordingly, we found that the genomes of the *Cucurbita* species contain two paralogs for each of the ethylene receptors ETR1, ERS1, and ETR2. No ERS2- or EIN4-like receptors were found in the genomes of these species. The *C. pepo* genome has two ETR1 duplicates (*CpETR1A* and *CpETR1B*), which showed more than 90% homology and mapped on chromosomes 7 and 11, and two ETR2 duplicates (*CpETR2A* and *CpETR2B*), which showed more than 89% homology and mapped on chromosome 8. The duplicates maintained the same molecular structure: six exons and five introns for the two *CpETR1* paralogs, and three exons and two introns for the two *CpETR2* paralogs (Fig. 5A, B).

The *etr1a* mutation was located at nucleotide position 284 of the first exon of the *CpETR1A* gene, and the *etr2b* mutation was located at nucleotide position 1018 of the first exon of the *CpETR2B* gene (Fig. 5A, B). The deduced ETR1A and ETR2B receptors had the same domains as those of the Arabidopsis homologs: an ethylene-binding domain with three transmembrane segments; one GAF domain; one histidine-kinase domain; and a response regulator receiver domain. The two mutations resulted in an amino acid substitution in each protein: A95V in the ethylene-binding domain of CpETR1A, and E340K in the coiled-coil domain between the GAF and histidine-kinase domains of CpETR2B (Fig. 5A, B). All ETR1- and ETR2-like proteins in the NCBI database were found to contain the WT amino acid at these two particular positions, indicating that the amino acids affected by the two mutations are highly conserved in very different plant species (Supplementary Figs S2 and S3).

To assess the genetic relationships between ETR1- and ETR2-like ethylene receptors in *C. pepo*, a phylogenetic tree was inferred from the ETR1, ERS1, and ETR2 protein sequences of different cucurbit species—*Cucurbita maxima*, *Cucurbita moschata*, *C. melo*, *C. sativus*, and *C. lanatus*—together with the five ethylene receptors of Arabidopsis (Fig. 5C, Supplementary Table S4). In the other cucurbits, only one single gene was found for each ethylene receptor. In agreement with the allotetraploid origin of the genus *Cucurbita* (Sun *et al.*, 2017), only one of the paralogs in subgenome B (*CpETR1B*, *CpERS1B*, and *CpETR2B* in *C. pepo*) clustered with their orthologs in the rest of the cucurbit species (Fig. 5C).

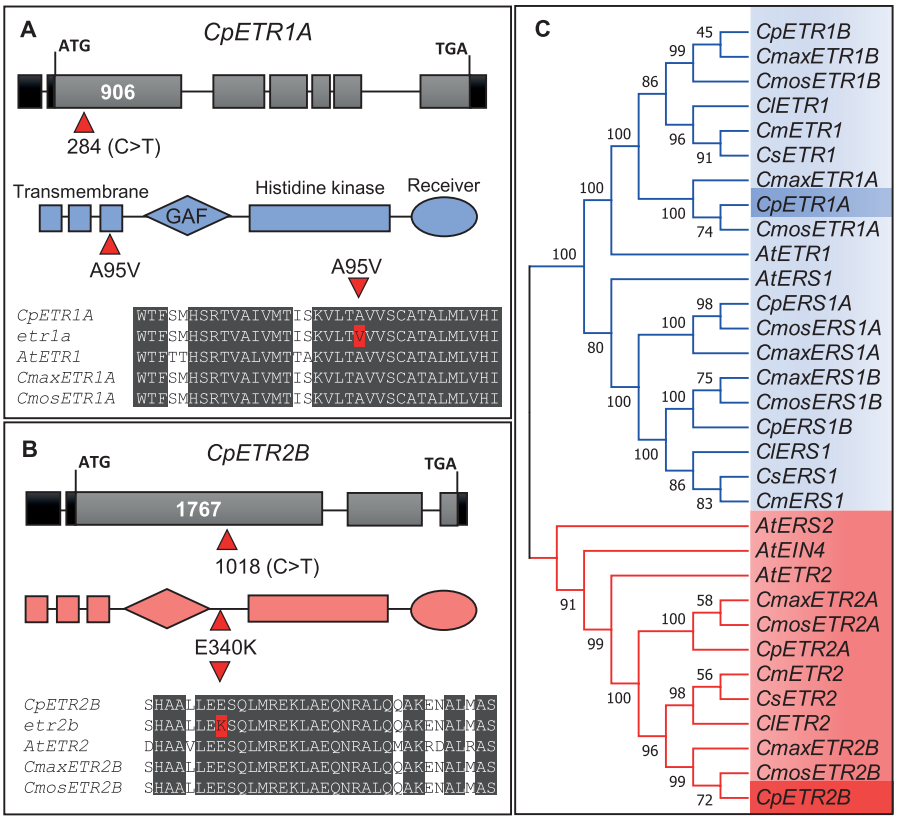


Fig. 5. Molecular structure of CpETR1A and CpETR2B, and phylogenetic analysis of ETR1- and ETR2-like ethylene receptors. (A) *CpETR1A*; (B) *CpETR2B*. Within each gene, black boxes indicate untranslated regions and grey boxes correspond to exons. Numbers indicate the size of the fragment (bp). Missense mutations in the gene, and amino acid substitutions in the protein, are indicated in red. EMS mutations C>T are numbered with respect to the coding sequence. The three transmembrane subdomains in the ethylene-binding domain, as well as the GAF, histidine-kinase, and receiver domains, are indicated for each ethylene receptor. (C) Phylogenetic relationships between the ethylene receptors of six cucurbit species (*Cucurbita pepo*, *Cucurbita moschata*, *Cucurbita maxima*, *Cucumis melo*, *Cucumis sativus*, and *Citrullus lanatus*) and those of Arabidopsis. The six ethylene receptors of *C. pepo* are positioned in three clusters corresponding to ETR1, ERS1, and ETR2 of Arabidopsis. In the three *Cucurbita* species, ETR1 and ETR2 are duplicated in relation to Arabidopsis.

Effects of etr1a *and* etr2b *on ethylene receptor gene expression*

The expression patterns of the mutated *CpETR1A and CpETR2B* and their duplicated paralogs (*CpETR1B* and *CpETR2A9*) were studied in leaves, roots, shoots, and shoot apices of WT plants (Fig. 6). The four genes were expressed in all the analyzed tissues, suggesting that both duplicates maintain their expression in these tissues.

We also investigated whether the *etr1a* or *etr2b* mutations alter the patterns of expression of the four *ETR* genes. Gene expression was compared in WT and mutant female floral buds at very early stages of development, that is, when stamen arrest takes place in the WT female flower (stage T0), and at 1 day before anthesis of WT flowers (stage T5); in the latter case, we separated the flower into the ovary and a tissue comprising the petals, style, and stigma (Fig. 7). The *etr1a* mutation inhibited the expression of *CpETR1A* and *CpETR2A* in the tissue that comprised the petals, style, and stigma of T5 flowers, and also reduced the transcription of *CpETR2B* in T0 flowers (Fig. 7). By contrast, the *etr2b* mutation inhibited the expression of *CpETR1B* and *CpETR2A* in T0 floral buds, and the expression of *CpETR1A* in the ovary of T5 flowers.

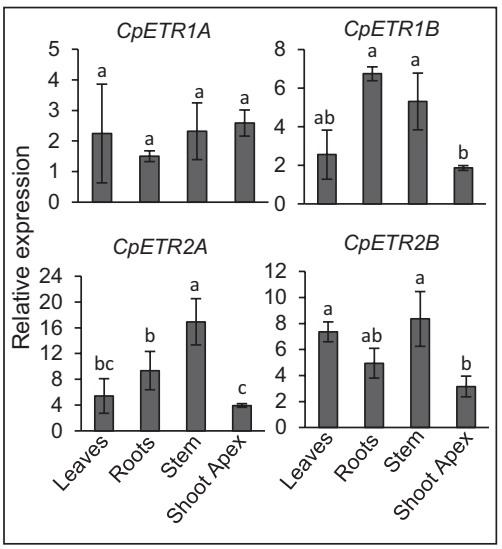


Fig. 6. Expression patterns of *CpETR1* and *CpETR2* ethylene receptor genes in different plant organs of *C. pepo* WT plants. The relative level of each transcript was quantified by qRT–PCR in three independent replicates of each tissue. The shoot apex consists not solely of the apical meristem but also of small leaves and floral buds. Different letters indicate statistically significant differences ($P<0.05$) between samples.

Discussion

etr1a *and* etr2b *are two missense mutations in ethylene receptors leading to semi-dominant ethylene insensitivity*

Ethylene is perceived by a family of two-component histidine kinase receptors that repress the ethylene signaling cascade in the absence of ethylene but become inactivated upon ethylene binding (Hua and Meyerowitz, 1998). Mutations in ethylene receptor genes fall into two main categories: (i) dominant gain-of-function mutations, conferring ethylene insensitivity, and (ii) recessive loss-of-function mutations that have little effect as single mutations but show a constitutive ethylene response in combination, for example, in double, triple, and quadruple mutants of Arabidopsis. Both *etr1a* and *etr2b* of *C. pepo* correspond to the first type of mutation, since they are semi-dominant and result in plant ethylene insensitivity (Fig. 1).

In Arabidopsis dominant mutants, the ethylene-insensitive phenotypes are caused by single amino acid substitutions in the transmembrane ethylene-binding domain of any of the five ethylene receptors described in this species (Bleecker *et al.*, 1988; Chang *et al.*, 1993; Guzmán and Ecker, 1990; Hua *et al.*, 1995, 1998; Wang *et al.*, 2006). The *etr1a* mutation described here is also a missense mutation (A95V), situated in the third transmembrane domain of the N-terminal ethylene-binding site of *CpETR1A* (Fig. 5), which causes a strong reduction in ethylene sensitivity in etiolated seedlings (Fig. 1). The mutation is contained in a conserved segment of the protein, close to the T94M mutation of Arabidopsis ETR1, which is known to disrupt the ability of the receptor to bind ethylene and to strongly affect ethylene sensitivity (Wang *et al.*, 2006; Resnick *et al.*, 2008). The *etr2b* mutation, however, is a missense mutation (E340K) within the coiled-coil domain between the GAF and histidine-kinase domains of *CpETR2B* (Fig. 5). Given that this domain does not participate in ethylene binding, it is likely that the ethylene insensitivity of *etr2b* is caused by a lack of transduction of the ethylene signal, as has been suggested for other dominant ethylene-insensitive mutations (Hall *et al.*, 1999). Thus, *etr1a* may disrupt the ethylene-binding site, and *etr2b* may alter ethylene signal transduction, but both mutations

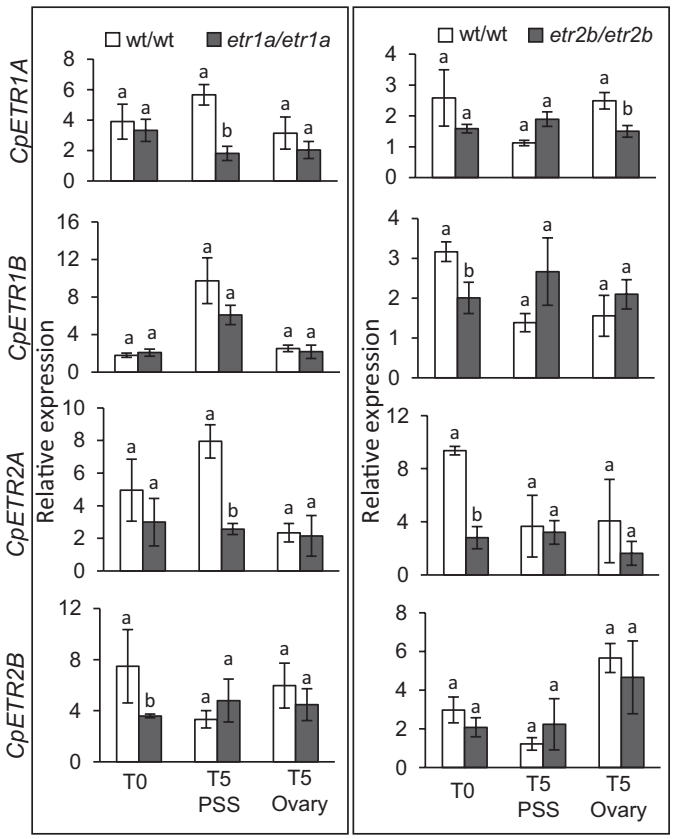


Fig. 7. Relative expression of *CpETR1* and *CpETR2* ethylene receptor genes in female flowers of *C. pepo*. The relative level of each transcript was quantified by qRT–PCR in three independent replicates of each tissue. T0 corresponds to complete female flowers 2 mm in length; T5 corresponds to pre-anthesis-stage female flowers, separated into the ovary and a tissue comprising the petals, style, and stigma (PSS). The comparison of gene expression was performed between homozygous WT and homozygous mutant flowers derived from plants in the same segregating population. Different letters indicate statistically significant differences ($P<0.05$) between samples.

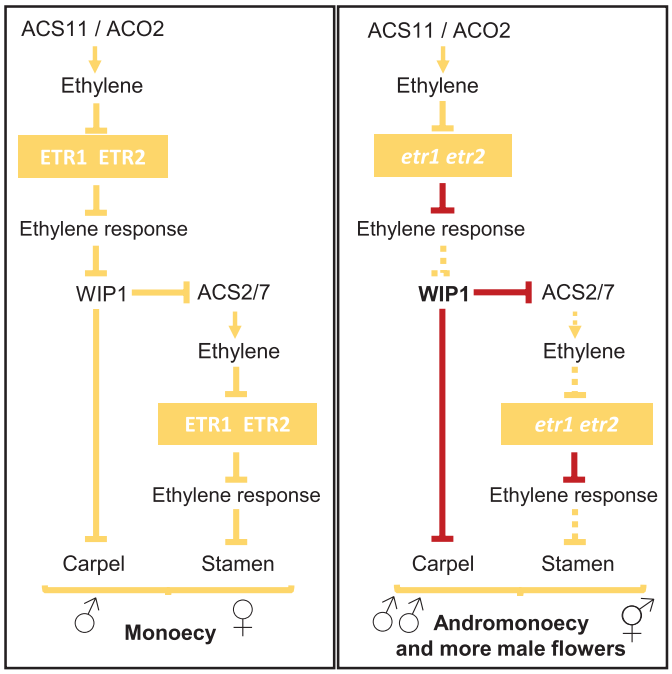


Fig. 8. Model of sex determination in cucurbit species, integrating the function of the ethylene receptors ETR1 and ETR2 of *C. pepo* with other sex-determining genes identified in *C. melo* and *C. sativus* (*ACS11, ACO2, WIP1*, and *ACS2/7*). The ethylene biosynthesis enzymes ACS11 and ACS2/7 have different spatiotemporal expression patterns, causing the arrest of carpels or stamens required for a flower to develop as male or female, respectively (left panel). The two biosynthetic pathways are connected by the transcription factor WIP1, which represses the transcription of ACS2/7, and is negatively regulated by the ethylene-producing enzymes ACS11and ACO2 (Boualem *et al.*, 2015; Che and Zhang, 2019). The ethylene receptors ETR1 and ETR2 should perceive and transmit signaling of the ethylene synthesized by both the ACS11 and ACS2/7 pathways. The inhibition of the ACS2/7 ethylene response releases the arrest of stamens in female flowers, resulting in the production of bisexual or hermaphrodite flowers (andromonoecy). On the other hand, the inhibition of the ACS11 ethylene response can induce WIP1, which enhances the arrest of carpels and so leads to the formation of male flowers. Red and dotted lines in the right panel indicate increased or decreased effects, respectively, produced by the *etr1* and *etr2* mutations.

should convert the CpETR1A or CpETR2B receptors to a constitutive signaling-on state that represses the ethylene response (Fig. 8). The differing levels of ethylene insensitivity shown by single *etr1a* and *etr2b* mutants of *C. pepo* could indicate that *CpETR1A* (subfamily I) has a more prominent role in ethylene perception than *CpETR2B* (subfamily II). In Arabidopsis, both *etr1-1* and *etr1-4* mutants are insensitive to ethylene, but *etr1-2*, *etr2-1*, and *ein4-3* maintain a reduced response to ethylene (Hall *et al.*, 1999).

A whole-genome duplication occurred just before speciation of the genus *Cucurbita,* which explains why only the species of the genus *Cucurbita* have two subgenomes (A and B) (Sun *et al.*, 2017; Montero-Pau *et al.*, 2018). *In C. pepo*, we found that the paralogs of *CpETR1* (*CpETR1A* and *CpETR1B*) and *CpETR2* (*CpETR2A* and *CpETR2B*) in each subgenome are both expressed and could be functional (Figs 6 and 7). This tetraploid origin of *C. pepo* could account for the semi-dominant nature of *etr1a* and *etr2*b. In Arabidopsis, ethylene-insensitive mutants show the same phenotype in both homozygous and heterozygous conditions; however, an extra WT allele of the same gene in triploid plants (*mut*/wt/wt) for either *etr1-1*, *etr1-2*, or *ein4-3* reduced ethylene insensitivity compared with homozygous or heterozygous diploid mutants (*mut*/*mut* or wt/*mut*). This suggests that the dominant insensitivity of these mutant alleles is mediated by interaction between WT and mutated ethylene receptor isoforms (Hall *et al.*, 1999). Our findings also indicate that the CpETR1A and CpETR2B ethylene receptors exert their action cooperatively rather than independently. Different members of the Arabidopsis ethylene receptors form homomeric and heteromeric complexes that may facilitate receptor signal output (Binder and Bleecker, 2003; Xie *et al.*, 2006; Grefen *et al.*, 2008; Gao *et al.*, 2008; Gao and Schaller, 2009), and mutant receptors can repress the ethylene response only when coupled with a WT receptor, which implies that each tissue can have mixed receptor complexes with different receptor signal output strengths (Li *et al.*, 2009).

The cooperation between WT and mutant alleles of *CpETR1A* and *CpETR2B* is also suggested by the phenotypes of single and double mutants, where homozygous single mutants for *etr1a* or *etr2b* (85% and 63.7% ethylene insensitivity, respectively) show similar ethylene insensitivity to the heterozygous double mutant wt/*etr1* wt/*etr2b* (61.3%) but stronger insensitivity than that of the heterozygous single mutants wt/*etr1a* and wt/*etr2b* (53.5% and 43.5%, respectively). The suppression of the ethylene response in *C. pepo* is therefore dependent on the dosage of WT and mutant alleles for either *CpETR1A* or *CpETR2B* (Figs 8 and 9).

Ethylene insensitivity in the two *etr* mutants was not, however, associated with the expression level of the genes. The *etr1a* and *etr2b* mutations scarcely affected the expression of ethylene receptor genes, although some of them were down-regulated in female flowers, probably owing to a mechanism that aids the recovery of ethylene sensitivity that may be lost as a result of the mutations. Functional compensation between ethylene receptor gene families has been found to occur between *NR* and *LeETR4* of tomato (Tieman *et al.*, 2000) and in certain loss-of-function ethylene receptor mutants of Arabidopsis (Zhao *et al.*, 2002; Chen *et al.*, 2007; Harkey *et al.*, 2018), but not in rice (Wuriyanghan *et al.*, 2009). Nevertheless, given that some of the *ETR* genes were not down-regulated in all *etr1a* or *etr2b* tissues, it is also likely that reduction in gene expression reflects a differential regulation of the analyzed ethylene receptor genes.

Mutations in CpETR1A *and* CpETR2B *alter sex determination and expression*

Sex determination in individual floral buds of monoecious and dioecious species is controlled by diverse mechanisms that suppress the development of either stamen or carpel primordia in floral buds that will result in female or male flowers, respectively (Fig. 8; Kater *et al.*, 2001; Bai *et al.*, 2004). In the Cucurbitaceae, the main sex regulator is ethylene, but the genetic network controlling sex determination in these species is still poorly understood. With the exception of *WIP1*, all major genes controlling sex determination encode key enzymes involved in ethylene biosynthesis (ACS and ACO; Fig. 8). The gene *ACS2/ACS7* controls monoecy, and loss-of-function mutations lead to plants with male and hermaphrodite flowers (andromonoecy) (Boualem *et al.*, 2008; Boualem *et al.*, 2009; Martínez *et al.*, 2014; Ji *et al.*, 2016; Manzano *et al.*, 2016). Mutations in this gene do not affect sexual expression, that is, the ratio between male and female flowers in the plant (Martínez *et al.*, 2014; Manzano *et al.*, 2016). On the other hand, the genes *ACS11* and *ACO2* are required for the female flower development pathway, and mutations in these genes result in plants with only male flowers (androecy) (Boualem *et al.*, 2015; Chen *et al.*, 2016). Finally, the transcription factor WIP1 is required for male flower development and is negatively regulated by ACS11, and its dysfunction results in plants with only female flowers (gynoecy) (Boualem *et al.*, 2015; Hu *et al.*, 2017).

So far, no ethylene receptor has been positioned in the genetic network controlling sex determination in this group of species (Fig. 8), although there was some evidence indicating their participation. Thus, the transcription level of *ETR2* and *ERS1* is higher in gynoecious than monoecious apical shoots of *C. sativus* (Yamasaki *et al.*, 2000), and and the down-regulation of *CsETR1* in the stamens of female cucumber flowers appears to be required for the arrest of stamen development (Wang *et al.*, 2010). Moreover, transgenic *C. melo* plants overexpressing the Arabidopsis ethylene-insensitive allele *etr1-1* are altered in sex determination and sex expression (Little *et al.*, 2007; Switzenberg *et al.*, 2015). Given that *etr1a* and *etr2b* not only disrupt female flower development (converting monoecy into andromonoecy) but also significantly increase the number of male flowers in the plant, it is likely that ETR1 and ETR2 integrate the two ethylene biosynthesis pathways that result in the determination of male and female flowers, perceiving and signaling the ethylene produced by ACS2/7 as well as that produced by ACS11 and ACO2 (Fig. 8).

The mechanisms triggering the sex of each specific floral meristem must be regulated by the level of ethylene sensitivity conferred by ethylene receptors (Fig. 9). In fact, the degree of conversion of female flowers into bisexual and hermaphrodite flowers, the delay in pistillate flowering transition, and the increase in the number of male flowers per plant are all correlated with the level of ethylene sensitivity in *etr1a* and *etr2b* single and double mutants. The final level of ethylene sensitivity in the tissue will be affected by the strength of each mutation, but

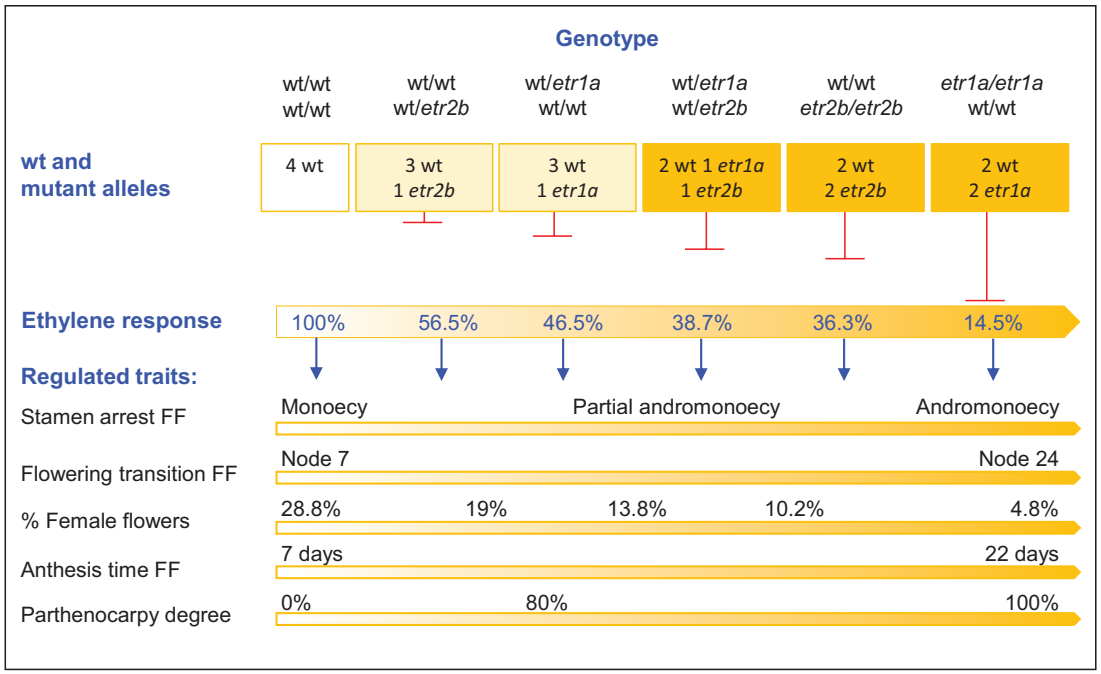


Fig. 9. Cooperation of the CpETR1A and CpETR2B ethylene receptors in the regulation of the ethylene response and ethylene-associated traits in *C. pepo* flower and fruit development. In the absence of ethylene, CpETR1A and CpETR2B repress the ethylene response. In the presence of ethylene, the repression of the ethylene response is determined by the dosage of WT and mutant alleles, and also by the strength of each mutant allele, with *etr1a* having a stronger effect than *etr2b*. The resulting ethylene sensitivity regulates the development of pistillate floral organs and sex determination. As the number of mutant alleles increases, the response to ethylene is reduced. This increases the degree of andromonoecy, delays the pistillate (female) floral transition, and decreases the number of pistillate flowers per plant. It also delays the maturity of pistillate flowers, favoring growth of the ovary and parthenocarpic development of the fruit. FF, female flowers.

also by the number of WT and mutant ethylene receptors in the tissue (Hall *et al.*,1999), and by the cooperation between them for repressing ethylene signaling (Fig. 9).

The ethylene-insensitive mutations also alter the growth rate of petals and carpels, making mutant bisexual and hermaphrodite flowers reach anthesis later than WT female flowers. This delay in anthesis time could be associated with the development of stamens, since the anthesis time of the mutant hermaphrodite flower was similar to that of the male flower. Moreover, the mutant ovary continues to grow as long as the petals remain green and do not reach anthesis (Fig. 4). These data suggest that fruit set in *C. pepo* is not triggered by pollination, but is a default developmental program, which has a checking point at anthesis. If the flower is pollinated, development of the ovary continues and the fruit sets. In the absence of pollination and fertilization, growth of the ovary is aborted. Fruit set is known to be regulated positively by hormones such as auxins and gibberellins, and negatively by ethylene (Martínez *et al.*, 2013; Shinozaki *et al.*, 2018; Shnaider *et al.*, 2018). Therefore, the reduction of ethylene sensitivity in the *etr1a* and *etr2b* mutants could trigger parthenocarpy by delaying flower anthesis and consequently the checking point for fruit set.

Supplementary data

Supplementary data are available at *JXB* online.

Table S1. Sequence and coverage statistics.

Table S2. Primers and TaqMan probes used for genotyping *etr1a* and *etr2b* mutations.

Table S3. Primers for qRT–PCR analysis.

Table S4. Proteins used to perform the phylogenetic analysis.

Table S5. Segregation of ethylene-sensitive, -intermediate and -insensitive plants in the offspring of backcrossed (BC_1 and BC_2) and selfed (BC_1S_1 and BC_2S_1) generations.

Fig. S1. Distribution of staminate and pistillate flowers in the 40 first nodes of the plant.

Fig. S2. Alignment of the *CpETR1A* amino acid sequence with homologous sequences from diverse species.

Fig. S3. Alignment of the *CpETR2B* amino acid sequence with homologous sequences from diverse species.

Acknowledgements

This work was supported by grants AGL2014-54598-C2-1-R and AGL2017-82885-C2-1-R, which were funded partly by the European Regional Development Fund and partly by the Spanish Ministry of Science and Innovation, and grant P12-AGR-1423, funded by Junta de Andalucía, Spain.

Conflict of interest

The authors state that no conflict of interest exists regarding this publication.

References

Bai SL, Peng YB, Cui JX, Gu HT, Xu LY, Li YQ, Xu ZH, Bai SN. 2004. Developmental analyses reveal early arrests of the spore-bearing parts of

reproductive organs in unisexual flowers of cucumber (*Cucumis sativus* L.). Planta **220**, 230–240.

Binder BM, Bleecker AB. 2003. A model for ethylene receptor function and 1-methylcyclopropene action. Acta Horticulturae **628**, 177–187.

Bleecker AB, Estelle MA, Somerville C, Kende H. 1988. Insensitivity to ethylene conferred by a dominant mutation in *Arabidopsis thaliana*. Science **241**, 1086–1089.

Boualem A, Fergany M, Fernandez R, *et al*. 2008. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science **321**, 836–838.

Boualem A, Troadec C, Camps C, *et al*. 2015. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. Science **350**, 688–691.

Boualem A, Troadec C, Kovalski I, Sari MA, Perl-Treves R, Bendahmane A. 2009. A conserved ethylene biosynthesis enzyme leads to andromonoecy in two *Cucumis* species. PLoS ONE **4**, e6144.

Byers RE, Baker LR, Sell HM, Herner RC, Dilley DR. 1972. Ethylene: a natural regulator of sex expression of *Cucumis melo* L. Proceedings of the National Academy of Sciences, USA **69**, 717–20.

Chang C, Kwok SF, Bleecker AB, Meyerowitz EM. 1993. *Arabidopsis* ethylene-response gene *ETR1*: similarity of product to two-component regulators. Science **262**, 539–544.

Che G, Zhang X. 2019. Molecular basis of cucumber fruit domestication. Current Opinion in Plant Biology **47**, 38–46.

Chen H, Sun J, Li S, *et al*. 2016. An ACC oxidase gene essential for cucumber carpel development. Molecular Plant **9**, 1315–1327.

Chen YF, Shakeel SN, Bowers J, Zhao XC, Etheridge N, Schaller GE. 2007. Ligand-induced degradation of the ethylene receptor ETR2 through a proteasome-dependent pathway in *Arabidopsis*. Journal of Biological Chemistry **282**, 24752–24758.

Dellaporta SL, Calderon-Urrea A. 1993. Sex determination in flowering plants. The Plant Cell **5**, 1241–1251.

Den Nijs APM, Visser DL. 1980. Induction of male flowering in gynoecious cucumbers (*Cucumis sativus* L.) by silver ions. Euphytica **29**, 273–280.

Gao Z, Schaller GE. 2009. The role of receptor interactions in regulating ethylene signal transduction. Plant Signaling & Behavior **4**, 1152–1153.

Gao Z, Wen CK, Binder BM, Chen YF, Chang J, Chiang YH, Kerris RJ 3rd, Chang C, Schaller GE. 2008. Heteromeric interactions among ethylene receptors mediate signaling in *Arabidopsis*. Journal of Biological Cchemistry **283**, 23801–23810.

García A, Aguado E, Parra G, *et al*. 2018. Phenomic and genomic characterization of a mutant platform in *Cucurbita pepo*. Frontiers in Plant Science **9**, 1049.

Grefen C, Städele K, Růzicka K, Obrdlik P, Harter K, Horák J. 2008. Subcellular localization and in vivo interactions of the *Arabidopsis thaliana* ethylene receptor family members. Molecular Plant **1**, 308–320.

Guzmán P, Ecker JR. 1990. Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. The Plant Cell **2**, 513–523.

Hall AE, Chen QG, Findell JL, Schaller GE, Bleecker AB. 1999. The relationship between ethylene binding and dominant insensitivity conferred by mutant forms of the ETR1 ethylene receptor. Plant Physiology **121**, 291–300.

Harkey AF, Watkins JM, Olex AL, DiNapoli KT, Lewis DR, Fetrow JS, Binder BM, Muday GK. 2018. Identification of transcriptional and receptor networks that control root responses to ethylene. Plant Physiology **176**, 2095–2118.

Hu B, Li D, Liu X, Qi J, Gao D, Zhao S, Huang S, Sun J, Yang L. 2017. Engineering non-transgenic gynoecious cucumber using an improved transformation protocol and optimized CRISPR/Cas9 system. Molecular Plant **10**, 1575–1578.

Hua J, Chang C, Sun Q, Meyerowitz EM. 1995. Ethylene insensitivity conferred by *Arabidopsis ERS* gene. Science **269**, 1712–1714.

Hua J, Meyerowitz EM. 1998. Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*. Cell **94**, 261–271.

Hua J, Sakai H, Nourizadeh S, Chen QG, Bleecker AB, Ecker JR, Meyerowitz EM. 1998. *EIN4* and *ERS2* are members of the putative ethylene receptor gene family in Arabidopsis. The Plant Cell **10**, 1321–1332.

Jamilena M, Mariotti B, Manzano S. 2008. Plant sex chromosomes: molecular structure and function. Cytogenetic and Genome Research **120**, 255–264.

Ji G, Zhang J, Zhang H, *et al*. 2016. Mutation in the gene encoding *1-aminocyclopropane-1-carboxylate synthase 4 (CitACS4)* led to andromonoecy in watermelon. Journal of Integrative Plant Biology **58**, 762–765.

Kater MM, Franken J, Carney KJ, Colombo L, Angenent GC. 2001. Sex determination in the monoecious species cucumber is confined to specific floral whorls. The Plant Cell **13**, 481–493.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution **33**, 1870–1874.

Li Z, Huang S, Liu S, *et al*. 2009. Molecular isolation of the M gene suggests that a conserved-residue conversion induces the formation of bisexual flowers in cucumber plants. Genetics **182**, 1381–1385.

Little HA, Papadopoulou E, Hammar SA, Grumet R. 2007. The influence of ethylene perception on sex expression in melon (*Cucumis melo* L.) as assessed by expression of the mutant ethylene receptor, *At-etr1-1*, under the control of constitutive and floral targeted promoters. Sexual Plant Reproduction **20**, 123–136.

Malepszy S, Niemirowicz-Szczytt K. 1991. Sex determination in cucumber (*Cucumis sativus*) as a model system for molecular biology. Plant Science **80**, 39–47.

Manzano S, Aguado E, Martínez C, Megías Z, García A, Jamilena M. 2016. The ethylene biosynthesis gene *CitACS4* regulates monoecy/andromonoecy in watermelon (*Citrullus lanatus*). PLOS ONE **11**, e0154362.

Manzano S, Martínez C, García JM, Megías Z, Jamilena M. 2014. Involvement of ethylene in sex expression and female flower development in watermelon (*Citrullus lanatus*). Plant Physiology and Biochemistry **85**, 96–104.

Manzano S, Martínez C, Megías Z, Gómez P, Garrido D, Jamilena M. 2011. The role of ethylene and brassinosteroids in the control of sex expression and flower development in *Cucurbita pepo*. Plant Growth Regulation **65**, 213–221.

Marco-Sola S, Sammeth M, Guigó R, Ribeca P. 2012. The GEM mapper: fast, accurate and versatile alignment by filtration. Nature Methods **9**, 1185–1188.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A. 2009. A transposon-induced epigenetic change leads to sex determination in melon. Nature **461**, 1135–1138.

Martínez C, Manzano S, Megías Z, Barrera A, Boualem A, Garrido D, Bendahmane A, Jamilena M. 2014. Molecular and functional characterization of *CpACS27A* gene reveals its involvement in monoecy instability and other associated traits in squash (*Cucurbita pepo* L.). Planta **239**, 1201–1215.

Martínez C, Manzano S, Megías Z, Garrido D, Picó B, Jamilena M. 2013. Involvement of ethylene biosynthesis and signalling in fruit set and early fruit development in zucchini squash (*Cucurbita pepo* L.). BMC Plant Biology **13**, 139.

McKenna A, Hanna M, Banks E, *et al*. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20**, 1297–1303.

Montero-Pau J, Blanca J, Bombarely A, *et al*. 2018. *De novo* assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus. Plant Biotechnology Journal **16**, 1161–1171.

Pannell JR. 2017. Plant sex determination. Current Biology **27**, R191–R197.

Perl-Treves R. 2004. Male to female conversion along the cucumber shoot: approaches to studying sex genes and floral development in Cucumis sativus. In: Ainsworth CC, ed. Sex determination in plants. London: Garland Science, 193–221.

Resnick JS, Rivarola M, Chang C. 2008. Involvement of *RTE1* in conformational changes promoting ETR1 ethylene receptor signaling in Arabidopsis. The Plant Journal **56**, 423–431.

Rudich J, Halevy AH, Kedar N. 1969. Increase in femaleness of three cucurbits by treatment with Ethrel, an ethylene releasing compound. Planta **86**, 69–76.

Shinozaki Y, Nicolas P, Fernandez-Pozo N, *et al*. 2018. High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. Nature Communications **9**, 364.

Shnaider Y, Mitra D, Miller G, Baniel A, Doniger T, Kuhalskaya A, Scossa F, Fernie AR, Brotman Y, Perl-Treves R. 2018. Cucumber ovaries inhibited by dominant fruit express a dynamic developmental program, distinct from either senescence-determined or fruit-setting ovaries. The Plant Journal **96**, 651–669.

Sun H, Wu S, Zhang G, *et al*. 2017. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. Molecular Plant **10**, 1293–1306.

Switzenberg JA, Beaudry RM, Grumet R. 2015. Effect of *CRC::etr1-1* transgene expression on ethylene production, sex expression, fruit set and fruit ripening in transgenic melon (*Cucumis melo* L.). Transgenic Research **24**, 497–507.

Tieman DM, Taylor MG, Ciardi JA, Klee HJ. 2000. The tomato ethylene receptors *NR* and *LeETR4* are negative regulators of ethylene response and exhibit functional compensation within a multigene family. Proceedings of the National Academy of Sciences, USA **97**, 5663–5668.

Wang W, Esch JJ, Shiu SH, Agula H, Binder BM, Chang C, Patterson SE, Bleecker AB. 2006. Identification of important regions for ethylene binding and signaling in the transmembrane domain of the ETR1 ethylene receptor of Arabidopsis. The Plant Cell **18**, 3429–3442.

Wang DH, Li F, Duan QH, Han T, Xu ZH, Bai SN. 2010. Ethylene perception is involved in female cucumber flower development. The Plant Journal **61**, 862–872.

Wuriyanghan H, Zhang B, Cao WH, *et al*. 2009. The ethylene receptor ETR2 delays floral transition and affects starch accumulation in rice. The Plant Cell **21**, 1473–1494.

Xie F, Liu Q, Wen CK. 2006. Receptor signal output mediated by the ETR1 N terminus is primarily subfamily I receptor dependent. Plant Physiology **142**, 492–508.

Yamasaki S, Fujii N, Takahashi H. 2000. The ethylene-regulated expression of *Cs-ETR2* and *Cs-ERS* genes in cucumber plants and their possible involvement with sex expression in flowers. Plant and Cell Physiology **41**, 608–616.

Zhang J, Shi J, Ji G, Zhang H, Gong G, Guo S, Ren Y, Fan J, Tian S, Xu Y. 2017. Modulation of sex expression in four forms of watermelon by gibberellin, ethephone and silver nitrate. Horticultural Plant Journal **3**, 91–100.

Zhao XC, Qu X, Mathews DE, Schaller GE. 2002. Effect of ethylene pathway mutations upon expression of the ethylene receptor ETR1 from Arabidopsis. Plant Physiology **130**, 1983–1991.

Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, eds. Evolving genes and proteins. New York: Academic Press, 97–166.