

EL ESTUDIO GENÓMICO DEL CACAO (*Theobroma cacao* L.); BREVE RECOPIACIÓN DE SUS BASES CONCEPTUALES

THE GENOMIC STUDY OF CACAO (*Theobroma cacao* L.): BRIEF RECOPIATION OF ITS CONCEPTUAL BASES

Ricaño-Rodríguez, J.^{1*}; Ramos-Prado, J. M.¹; Cocolletzi-Vásquez, E.²; Hipólito-Romero, E.¹

¹Universidad Veracruzana, Centro de EcoAlfabetización y Diálogo de Saberes, Campus USBI, Col. Emiliano Zapata, Xalapa, Veracruz, México.

*Autor de Correspondencia: jricano@uv.mx

ABSTRACT

The cacao plant (*Theobroma cacao* L.) has been cultivated since more than 2,000 years ago by different Mesoamerican cultures. This species represents a significant biocultural resource for the world, due mostly to its importance in the production and commercialization of chocolate. Nowadays, the use and application of genomic tools make relatively possible the characterization of "genotypes" as well as the elucidation of genetic differences between hundreds of individuals at the population level. Current methods of molecular diagnosis, such as microarrangements, massive sequencing of new generation, assemblage and annotation of genomes and genetic expression, help generate data, as well as programming languages and mathematical algorithms that are used in biotechnological processes, whose objective is to improve the understanding and exploitation of the biology of this type of specimens.

Keywords: *Theobroma cacao*, biotechnology, genomics, bioinformatics.

RESUMEN

La planta de cacao (*Theobroma cacao* L.) ha sido cultivada desde hace más de 2,000 años por distintas culturas Mesoamericanas. Esta especie representa un recurso biocultural significativo para el mundo, debido principalmente a su importancia en la producción y comercialización del chocolate. Hoy en día, el uso y aplicación de herramientas genómicas facilitan relativamente la caracterización de "genotipos" al igual que la dilucidación de diferencias genéticas entre cientos de individuos a nivel poblacional. Métodos actuales de diagnóstico molecular, como microarreglos, secuenciaciones masivas de nueva generación, ensamblado y anotación de genomas y expresión génica, ayudan a la generación de datos, así como lenguajes de programación y algoritmos matemáticos que son utilizados en procesos biotecnológicos, cuya finalidad es mejorar el entendimiento y explotación de la biología de este tipo de especímenes.

Palabras clave: *Theobroma cacao*, biotecnología, genómica, bioinformática.

INTRODUCCIÓN

Theobroma cacao L., es una especie tropical de naturaleza diploide. Debido a que es explotada por alrededor de seis millones de agricultores en todo el mundo, y es un medio de subsistencia de más de 40 millones de personas, posee una importancia económica global considerable (Wickramasuriya y Dunwell, 2017; Beg *et al.*, 2017). Se cree que su cultivo se originó en la cuenca amazónica, y hoy en día es explotado en muchas regiones de los trópicos húmedos. La mayor parte de la producción mundial de cacao (a.c. 90%) proviene de pequeños agricultores. Actualmente se conocen más de 20 especies dentro del género *Theobroma* L., y *T. cacao*, (Figura 1) es el único linaje que se cultiva de manera extensiva (Hipólito-Romero *et al.*, 2017).

Esta especie se reagrupa en tres variedades genéticas que delimitan de manera respectiva sus características morfológicas y anatómicas: *i.e.* cacao criollo (*T. cacao* spp. *Criollo*), forastero (*T. cacao* spp. *Sphaerocarpum*) y trinitario. De éstas, la variedad criolla es bien conocida por su calidad organoléptica deseable y proporciona la materia prima con la que se producen los chocolates de sabor fino que representan hasta el 10% de la producción mundial. Hoy en día, la mayor parte de la producción total de chocolate (a.c. 80%) proviene del tipo de cacao forastero; tal variedad se ve favorecida sobre la criolla por su naturaleza resistente a ciertas enfermedades, además de su alto rendimiento. Por otra parte, los frutos son relativamente más baratos que los de tipo criollo. La tercera variedad genética; trinitaria, es un híbrido producido a partir de cruzamientos entre las variedades criollas y forasteras, y se cultiva en muchas partes de Centro y Sudamérica, África, Sudeste Asiático y Oceanía. Lo anterior debido a su aroma característico, buena productividad y resistencia a diversas enfermedades (Pridmore *et al.*, 2000).

El genoma del cacao es relativamente pequeño y se encuentra organizado en diez cromosomas; es decir, que su tamaño es aproximadamente del doble que el de *Arabidopsis thaliana*, un modelo ecotipo de dicotiledónea utilizado ampliamente

en estudios de genética aplicada. De acuerdo con Motamayor *et al.* (2013), el genoma de la variedad de cacao más cultivada en el mundo; el clon de *T. cacao* Matina 1-6, posee un tamaño aproximado de 445 Mpb (millones de pares de bases), el cual es considerablemente más grande que el genoma de la variedad criolla (430 Mpb) previamente descrita por Argout *et al.* (2011). Así, la información obtenida de este último genoma sugiere la presencia de 28,798 genes que transcriben más de 682 familias de proteínas en el cacao. Muchos de estos transcritos se encuentran relacionados con mecanismos de resistencia a enfermedades, biosíntesis de lípidos, rutas biosintéticas de flavonoides y terpenoides, y de manera general; cientos de compuestos elicitors que conjuntamente conforman el metaboloma de la especie.

A la luz de dichas consideraciones, distintas herramientas bioinformáticas actuales como; (1) búsqueda de alineamientos básicos locales y dominios conservados, (2) generación de bases de datos de familias de proteínas, (3) ensamblado y anotación de genomas, entre otras; llevan a cabo de manera conjunta la dilucidación de la variabilidad genética, filogenia evolutiva, así como transcriptoma, proteoma, secretoma y metaboloma de cientos de especies de interés agronómico.

Propuestas sobre el origen y diversidad genética del cacao

La diversidad genética del cacao ha sido estudiada a través de marcadores moleculares como: isoenzimas, ADN mitocondrial y de cloroplastos, RFLP; Polimorfismo de Longitud de Fragmentos de Restricción, al igual que RAPD; Amplificación Aleatoria de ADN Polimórfico, entre otros (Santana *et al.*, 2016). El principal interés en realizar este tipo de indagación se origina a partir de la necesidad de conocer la diferencia

entre las variedades criollas y forasteras. Por otra parte, un segundo punto de partida es el antecedente de que la mayor diversidad de especímenes de cacao se encuentra entre regiones adyacentes a Sudamérica y la cuenca amazónica (Whitkus *et al.*, 1998). La amplia diversidad genética de plantas de cacao de esta última región, aunado a sus notables características morfológicas y resistencia a



Figura 1. Árbol de cacao en etapa de desarrollo temprano (*Theobroma cacao* L.). Fotografía tomada en el rancho El Papayal, Mpio. de Nautla, Veracruz.

enfermedades, fortalecen la hipótesis de que dicho lugar es el centro de diversidad y posible origen de todas las especies (Chessman, 1944). Es así, que de manera previa ya se habían registrado ciertas diferencias entre las variedades criollas y forasteras, sin embargo, debido a las aparentes comparaciones tempranas entre ambos tipos de frutos, Cuatrecasas (1964) fortaleció la hipótesis de una posible diferenciación natural de especímenes entre ambas regiones. Los estudios antes mencionados implicaron el registro de características de plantas nativas.

De acuerdo con Whitkus et al. (1998), las poblaciones de cacao representan un único segmento de diversidad genética, ya que al menos las poblaciones del sureste de México se componen de 12 alelos únicos agrupados que son significativamente diferentes de otros tipos de cacao. Respecto al origen y distribución de las variedades domesticadas, actualmente existen dos hipótesis que sugieren su diversidad poblacional. La primera de ellas; es que las poblaciones nativas de cacao existieron exclusivamente en América del sur. La variedad criolla predecesora fue transportada por los humanos desde la cuenca alta del Amazonas hasta Mesoamérica, donde fue eventualmente domesticada por los Mayas, o bien, por sus antecesores hace más de dos mil años (Motamayor et al., 2002). Por otra parte, como ya fue mencionado; Cuatrecasas (1964) sugiere que el cacao tuvo una distribución geográfica natural desde la región amazónica hasta el sureste de México como es común en muchos géneros de árboles neotropicales.

Motamayor et al. (2002) propusieron que el cacao criollo ancestral representa el grupo original de este tipo de cultivos, aunque no representa una subespecie separada que probablemente fue originada a partir de algunos individuos en América del sur y, que en un futuro pudieron ser dispersadas por el hombre en América central. La clasificación más reciente del germoplasma del cacao realizada por Motamayor et al. (2008) agrupa 10 clústeres principales basándose en estudios de diferenciación de geografía y genética poblacional, siendo estos: Marañón, Curaray, Criollo, Iquitos, Nanay, Contamana, Amelonado, Purús y Nacional y Guiana. Finalmente, en relación con los cacaos mexicanos, Avendaño et al. (2010) describieron 20 varietales distintos, los cuales son: varietal Carmelo, YUC-I, Lagarto, AGG-26, CRIM-21, CRIM-22, CRIM-30, CRIM-36, CRIM-38, CHRIM-10, FRIM-20, CRIM-24, CRIM-44, CRIM-56, CRIM-88, CRIM-I05, Lacandón-17, Lacandón-28, Loxicha y TAJR. Para tal clasificación se emplearon 42 descriptores provenientes de frutos, hojas, flores y semillas, propuestos por el INIFAP ante la Unión Internacional para la Protección de Nuevas Variedades de Plantas.

Secuenciaciones genómicas masivas de siguiente generación y su importancia en el estudio del cacao

La versión actualizada del genoma de la variedad criolla de cacao cuenta con la anotación del 99% de sus genes anclados a sus 10 cromosomas (Chromosome-Anchored Genes). Lanzada en enero de 2017 en su segunda versión, ésta se encuentra disponible en el sitio electrónico "Cocoa Genome Hub" (<http://cocoa-genome-hub.southgreen.fr/>). La disponibilidad de genomas completos de distintas variedades de cacao (Argout et al., 2011; Motamayor et al., 2013) permiten la identificación y caracterización de

nuevos genes de interés para los biotecnólogos y, por consiguiente, para los productores. Así también, dicha tarea contribuye al desarrollo de marcadores moleculares para futuras selecciones asistidas de genes objetivo (target genes) (Lopes et al., 2011).

La liberación de secuencias del genoma del cacao también ha proporcionado un camino para la identificación rápida, así como la caracterización funcional y estructural de muchas familias de genes, a través de vías de simulación computacional (in silico) y análisis de expresión. Por ejemplo, recientemente se han identificado y caracterizado tres proteínas tipo-legumain (*i.e.* proteasa cisteína-lisosomal; EC 3.4.22.34) denominadas TcLEG3, TcLEG6 y TcLEG9 de manera respectiva, que desempeñan diversos papeles en la muerte celular programada, germinación y desarrollo de semillas. Dichas predicciones fueron logradas mediante análisis in silico, modelos tridimensionales y estudios de expresión génica (Santana et al., 2016). Por otra parte, el análisis extensivo de una familia de genes relacionados con la patogénesis (PR-Genes) en el genoma de *T. cacao*, ha permitido la identificación de un conjunto de transcritos hipotéticos que probablemente participan en la mediación de respuestas de defensa (R-Genes; genes tipo R) contra fitopatógenos (Figura 2) de interés comercial, como; *Phytophthora palmivora* (Maora et al., 2017) y *Colletotrichum theobromicola* (Fister et al., 2016a). Las matrices construidas a partir de bibliotecas sustractivas también se han utilizado en investigaciones de respuestas moleculares en el desarrollo de la enfermedad de la mazorca negra (Legavre et al., 2015).

En términos de objetivos de reproducción, se sabe que los genes tipo R pueden dividirse en dos categorías principales. La primera está asociada a la resistencia al estrés biótico y la segunda hace referencia a los organismos patógenos (Jones y Dangl, 2006). Por citar un ejemplo, algunos métodos de diagnóstico molecular como MALDI-TOF MS (i.e. Espectrometría de Masas por Desorción/Ionización Láser Asistida por Matriz) se aplican de manera regular para la identificación rápida de *Moniliophthora perniciosa*, *P. palmivora*, *Phytophthora capsici*, *Phytophthora citrophthora*, *Phytophthora heveae*, *Ceratocystis cacaofunesta*, *Cyanophora paradoxa* y *Ceratocystis fimbriata* (dos Santos et al., 2017). Otro ejemplo concerniente a los agentes infecciosos que afectan al cacao y la importancia de su estudio molecular; es el virus CSSV (Cacao Swollen-Shoot Virus) de doble cadena de ADN y que pertenece al género Badnavirus (Muller, 2016). Éste se transmite en gran medida por chinches harinosas y aunque se han realizado esfuerzos para erradicar el problema mediante prácticas culturales, aún no se ha tenido éxito considerable alguno.

Es importante mencionar que las colecciones más grandes de germoplasma de cacao se encuentran en el Banco Internacional de Genes de Cacao de Trinidad (International Cocoa Genebank, Trinidad) (ICG, T) y pueden ser consultadas en la dirección electrónica correspondiente: <https://sta.uwi.edu/cru/index.asp> (2400 accesiones). Asimismo, el Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) en Costa Rica (<https://www.catie.ac.cr/en/>), posee una colección cercana a 1146 accesiones al igual que el Centro Internacional de Cuarentena del Cacao, ubicado en la Universidad de Reading, Reino Unido (<http://www.icgd.reading.ac.uk/icqc/>).

La interacción biotrófica entre *Moniliophthora perniciosa* y *Theobroma cacao* vista desde un enfoque molecular

La enfermedad del cacao denominada escoba de bruja (Witches' Broom Disease), causada por el hongo hemibiotrófico *M. perniciosa*, es una de las más devastadoras enfermedades de *T. cacao*. En contraste con otras interacciones hemibiotróficas, la etapa biotrófica de este fitopatógeno dura varios meses y es responsable de los síntomas más distintivos de la enfermedad, que comprenden entre otros; notables cambios morfológicos drásticos en los brotes infectados. Mediante enfoques duales de RNA-seq; secuenciación de RNA, es posible realizar aproximaciones a respuestas transcriptómicas completas de manera simultánea tanto en el cacao



Figura 2. A) Fruto de cacao infectado por *Phytophthora* sp., B) Micelio de hongo fitopatógeno colonizando tejido de cacao interno. Fotografía tomada en el rancho El Papayal, Mpio. de Nautla, Veracruz.

como en *M. perniciosa* durante su interacción biotrófica. La infección de *M. perniciosa* desencadena la reprogramación metabólica masiva en los tejidos enfermos del cacao y, aunque aparentemente vigorosos, los brotes infectados se comportan como estructuras energéticamente demandantes que se caracterizan por la inducción de respuestas de defensa muchas veces ineficaces, y bajo una clara privación de carbono (Legavre et al., 2015). De acuerdo con Teixeira et al. (2014), la infección culmina con el establecimiento de un proceso de senescencia en el hospedador, que señala el final de la etapa biotrófica de *Moniliophthora*.

Es de esta manera que los estudios transcriptómicos de los organismos fitopatógenos han ayudado en gran medida a caracterizar las etapas de nutrición fúngica y estrategias de infección que a su vez permiten identificar efectores hipotéticos de virulencia (Jones y Dangl, 2006). Curiosamente, *M. perniciosa* desarrolla una infección parasitaria a largo plazo que orquesta cambios en el metabolismo de la planta para aumentar la disponibilidad de nutrientes solubles antes de la muerte de dicho hospedador. Todo lo anterior, tiene lugar gracias a la interacción biotrófica en la que destaca la expresión

de diversos transcritos como: citoquinina deshidrogenasa1, glicósido hidrolasa familia19, inhibidor de proteínas I3, thaumatina relacionada con patogenicidad, glutatión S-transferasa C-extremo terminal, lacasa, aquaporina, proteína no apical meristemática, desaturasa de ácidos grasos tipo1, proteína de enlace clorofilal A-B y reductasa ferredoxina, entre otras (Teixeira et al., 2014).

Repositorios disponibles para el estudio genómico/funcional del cacao

Las herramientas bioinformáticas de acceso libre al igual que las bases de datos electrónicas, contribuyen de manera notable al desarrollo de investigaciones enfocadas en profundizar en el campo de las “ómicas” (i.e. neologismo inglés que hace referencia al estudio de la genómica, proteómica, metabolómica y transcriptómica, entre otras) y de ese modo implementar mejoras en los cultivos.

En el Cuadro 1 se muestra una lista de algunas herramientas informáticas/bases de datos de libre acceso para el público en general, donde se puede consultar

información referente al cacao; ya sea desde estudios genómicos hasta las principales enfermedades que lo afectan. La última versión del genoma de cacao está disponible en: Cacao Genome Database (CGD, <http://www.cacaogenomedb.org/>), la cual fue desarrollada en colaboración con MARS, USDA; Marketing Analysis & Reporting Services, U. S. Department of Agriculture/ARS; Agriculture Research Service, IBM; International Business Machines, Instituto de Genómica de la Universidad de Clemson, PIPRA; Public Intellectual Property Resource for Agriculture, HudsonAlpha Instituto de Biotecnología, Centro Nacional de Recursos Genómicos, Universidad de Indiana y la Universidad del Estado de Washington—Main Lab Bioinformatics. Asimismo, existen herramientas para el análisis genómico relativamente fáciles de utilizar como es el caso de “BLAST”; Basic Local Alignment Tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) que permite encontrar regiones similares entre grupos de secuencias biológicas.

La herramienta “Gbrowse” (<http://gmod.org/wiki/Gbrowse>) es otro algoritmo que trabaja combinando bases de

Cuadro 1. Lista de algunos repositorios disponibles para el estudio genómico/funcional del cacao.

Fuente	URL
International Cocoa Germplasm Database (ICGD)	http://www.icgd.reading.ac.uk/index.php
Cocoa Genome Hub	http://cocoa-genome-hub.southgreen.fr/
CocoaGenDB	http://cocoagendb.cirad.fr/
Dicots PLAZA 3.0	http://bioinformatics.psb.ugent.be/plaza/versions/plaza_v3_dicots/organism/view/Theobroma+cacao
Ensembl Plants	http://plants.ensembl.org/Theobroma_cacao/Info/Index
Cacao Genome Database	http://www.cacaogenomedb.org/
CacaoNet—Global Network for Cacao Genetic Resources	https://sites.google.com/a/cgxchange.org/cacaonet/home/partners-of-cacaonet
CEMID—Cocoa EST Marker Information Database	http://riju.byethost31.com/cocoa/?ckattemp=1
Phytozome v11.0	https://phytozome.jgi.doe.gov/pz/portal.html!info?alias=Org_Tcacao
Witches' Broom Disease Transcriptome Atlas	http://bioinfo08.ibi.unicamp.br/wbdatlas/
GenBank NCBI	http://www.ncbi.nlm.nih.gov/genome/?term=cocoa
Global Consumption Database	http://datatopics.worldbank.org/consumption/
Decoding the Cocoa Genome	https://www.cirad.fr/en/our-research/research-results/2010/decoding-the-cocoa-genome
Integrated Taxonomic Information System	https://www.itis.gov/servlet/SingleRpt/SingleRpt
UniProtKB	http://www.uniprot.org/uniprot/?query=Cocoa&sort=score
Ancient Cacao Map	http://en.ancientcacao.com/
Biodiversity International	https://www.biodiversityinternational.org/news/detail/mapping-the-biodiversity-status-of-over-10000-ghanaian-cocoa-farms/
The American Phytopathological Society	https://www.apsnet.org/publications/commonnames/Pages/Cacao.aspx
National Center for Genome Resources	http://www.ncgr.org/
Fundación Cacao México	http://www.cacaomexico.org/

Nota: Direcciones de sitios electrónicos actualizados hasta abril de 2018.

datos y páginas web interactivas para la manipulación y despliegue de anotaciones de genomas (Wickramasuriya y Dunwell, 2017). Por último, pero no menos importante, un recurso también destacable en el estudio de la genómica funcional es "UniProt"; Universal Protein Resource (<http://www.uniprot.org/>), repositorio informático para la secuenciación y anotación de proteínas. De esta manera, Uniprot es el resultado de un proyecto que surgió de la colaboración entre EMBL-EBI; Instituto Europeo de Bioinformática (<https://www.ebi.ac.uk/>), SIB; Instituto Suizo de Bioinformática (<https://www.sib.swiss/>) y PIR; Protein Information Resource (<https://pir.georgetown.edu/>).

CONCLUSIONES

La generación masiva de datos genómicos ha incrementado de manera sobresaliente en las últimas dos décadas, gracias al avance de las tecnologías de la información y el desarrollo de algoritmos cada vez más precisos, que, aunado a la actual tecnología de secuenciación genética de siguiente generación, permiten dilucidar los fenotipos de miles de especies. El género *Theobroma* además de ser considerado un recurso biocultural de suma importancia, trasciende al formar parte de los modelos de estudio bioinformáticos recientes, los cuales complementan el conocimiento de la fisiología y genómica funcional de estos especímenes. En conjunto, una correcta aplicación biotecnológica y agronómica de estos datos generados, permitirá llevar a cabo la explotación de dicho recurso bajo un enfoque sostenible y en beneficio de la sociedad, además de contribuir al conocimiento teórico-práctico de su naturaleza, al igual que un manejo integrado de las plagas fitopatógenas que lo aquejan.

AGRADECIMIENTOS

Al Cuerpo Académico UV-CA-263 Manejo y Conservación de Recursos Bioculturales de la Universidad Veracruzana. A la empresa NESTLÉ a través del Plan Cacao NESTLÉ de México. A la Dirección General de Superación Universitaria del Gobierno de México por otorgar el apoyo de Incorporación de NPTC al Dr. Jorge Ricaño Rodríguez, a través del Programa para el Desarrollo Profesional Docente para el Tipo Superior (PRODEP).

LITERATURA CITADA

Avendaño-Arrazate C. H., Ogata-Aguilar N., Gallardo-Méndez R. A., Mendoza-López A., Aguirre-Medina J. F., Sandoval-Esquivel A. 2010. Cacao Diversidad en México. Publicación especial N° 1. Instituto Nacional de Investigaciones Forestales, Agrícolas

y Pecuarias. Centro de Investigación Regional Pacífico Sur. Campo Experimental Rosario Izapa. Tuxtla Chico, Chiapas, México, 86 p.

- Argout X., Salse J., Aury J. M., Guiltinan M. J., Droc G., Gouzy J., Allegre M. et al. 2011. The genome of *Theobroma cacao*. *Nature Genetics* 43: 101-108.
- Beg M. S., Ahmad S., Jan K., Bashir K. 2017. Status, supply chain and processing of Cocoa - a review. *Trends in Food Science and Technology* 66: 108-116.
- Cheesman E. E. 1944. Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture* 21: 144-159.
- Cuatrecasas J. 1964. Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Contributions from the United States National Herbarium* 35: 379-614.
- Fister A. S., Mejia L. C., Zhang Y., Herre E. A., Maximova S. N., Guiltinan M. J. 2016a. *Theobroma cacao* L. pathogenesis-related gene tandem array members show diverse expression dynamics in response to pathogen colonization. *BMC Genomics* 17, 363.
- Hipólito-Romero E., Carcaño-Montiel M. G., Ramos-Prado J. M., Vázquez-Cabañas E. A., López-Reyes L., Ricaño-Rodríguez J. 2017. Efecto de inoculantes bacterianos edáficos mixtos en el desarrollo temprano de cultivares mejorados de cacao (*Theobroma cacao* L.) en un sistema agroforestal tradicional del norte de Oaxaca, México. *Revista Argentina de Microbiología* 49(4): 356-365.
- Jones J. D., Dangl J. L. 2006. The plant immune system. *Nature* 444: 323-329.
- Legavre T., Ducamp M., Sabau X., Argout X., Fouet O., Dedieu F., Surujdeo-Maharaj S. et al. 2015. Identification of *Theobroma cacao* genes differentially expressed during *Phytophthora megakarya* infection. *Physiological and Molecular Plant Pathology* 92: 1-13.
- Lopes U. V., Monteiro W. R., Pires J. L., Clement D., Yamada M. M., Gramacho K. P. 2011. Cacao breeding in Bahia, Brazil: strategies and results. *Crop Breeding and Applied Biotechnology* 11: 73-81.
- Maora J., Liew E., Guest D. 2017. Limited morphological, physiological and genetic diversity of *Phytophthora palmivora* from cocoa in Papua New Guinea. *Plant Pathology* 66: 124-130.
- Motamayor J. C., Risterucci A. M., Lopez P. A., Ortiz C. F., Moreno A., Lanaud C. 2002. Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89: 380-386.
- Motamayor J. C., Lachenaud P., da Silva e Mota J. W., Loor R., Kuhn D. N., Brown J. S., Schnell R. J. 2008. Geographic and genetic population differentiation of the amazonian chocolate tree (*Theobroma cacao* L.). *PLoS ONE* 3(10): e3311.
- Motamayor J. C., Mockaitis K., Schmutz J., Haiminen N., Livingstone D., Cornejo O., Findley S. D. et al. 2013. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biology* 14: r53.
- Muller E. 2016. Cacao Swollen Shoot Virus (CSSV): history, biology, and genome, en: *Cacao Diseases*. Eds Bailey B. A. and Meinhardt L. W. Switzerland Springer International Publishing. 358 p.
- Pridmore R. D., Cruzillat D., Walker C., Foley S., Zink R., Zwahlen M. C., Brussow H. 2000. Genomics, molecular genetics and the food industry. *Journal of Biotechnology* 78: 251-258.

- Santana J. O., Freire L., de Sousa A. O., Fontes Soares V. L., Gramacho K. P., Pirovani C. P. 2016. Characterization of the legumains encoded by the genome of *Theobroma cacao* L. *Plant Physiology and Biochemistry* 98: 162-170.
- dos Santos F. N., Tata A., Belaz K. R. A., Magalhaes D. M. A., Luz E. D. M. N., Eberlin M. N. 2017. Major phytopathogens and strains from cocoa (*Theobroma cacao* L.) are differentiated by MALDI-MS lipid and/or peptide/protein profiles. *Analytical and Bioanalytical Chemistry* 409: 1765-1777.
- Teixeira P. J., Thomazella D. P., Reis O., do Prado P. F., do Rio M. C., Fiorin G. L., José J., Costa G. G., Negri V. A., Mondego J. M., Mieczkowski P., Pereira G. A. 2014. High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. *Plant Cell*. 11: 4245-69.
- Whitkus R., de la Cruz M., Mota-Bravo L., Gómez-Pompa A. 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern México. *Theoretical and Applied Genetics* 96: 621-627.
- Wickramasuriya A. M., Dunwell J. M. 2017. Cacao biotechnology: current status and future prospects. *Plant Biotechnology Journal* 1: 1-14.

