

Informe administrativo sobre el uso de recursos en la supercomputadora Miztli

Título

Caracterización de la composición estructural de las maquinarias de producción viral de rotavirus y adenovirus. Proyecto de Investigación Regular. Responsable Técnico: Dr. Adán Oswaldo Guerrero Cárdenas. Laboratorio Nacional de Microscopía Avanzada. Instituto de Biotecnología. UNAM.

Resumen

Muchos virus se forman mediante el ensamblaje de estructuras proteicas altamente organizadas cuya biogénesis inicia en el interior de las células: algunos se comienzan a ensamblar en el citoplasma, como es el caso de los viroplasmatas (VP) formados por rotavirus, y otros en el núcleo como es el caso de los centros de replicación (CR) viral de adenovirus. Estas estructuras nanoscópica no se pueden observar con el microscopio óptico debido que la difracción de la luz que impide resolver estructuras inferiores a 250 nanómetros.

En el 2015-16 recibimos apoyo por parte de la Unidad de súper-cómputo de la UNAM para caracterizar estas maquinarias de producción viral mediante el uso de la microscopía de súper-resolución, esto empleando un protocolo de paralelización del análisis Bayesiano del parpadeo y foto-blanqueo (algoritmo 3B) que desarrollamos en el Laboratorio Nacional de Microscopía Avanzada de la UNAM (LNMA). En el periodo 2015-16 obtuvimos el primer mapeo de la distribución espacial de algunas de las proteínas que conforman 1) a los VP y 2) a los CR, ambos a una resolución inferior a los 50 nanómetros. En el periodo 2016-17 ampliamos el conocimiento sobre la estructura y función de los VP y los CR al caracterizar con mayor detalle ambas maquinarias de producción viral. Se enriqueció el conocimiento sobre la estructura de los VP mediante la caracterización espacial diversas proteínas que no incorporamos en el estudio del 2015, entre ellas: NSP2, NSP4, NSP5, VP4, VP6, y VP7. Así mismo, incluimos estudios sobre la localización del RNA de doble cadena de rotavirus dentro de los VP, así como su distribución relativa al retículo endoplásmico.

El 3 de marzo del 2016 nos aprobaron 499968 horas CPU. Consumimos solo un 28 % (140259.95 hCPU) de los recursos aprobados para concluir con los objetivos planteados en la propuesta original. Lo anterior se debió a que migramos nuestros protocolos de análisis por microscopía de súper-resolución (cómputo paralelo en CPU), al uso de tarjetas gráficas, GPUs.

El presente documento constituye el final de nuestras actividades en la MIZTLI.

Avances (1 cuartilla).

Dada las limitaciones de espacio, en la presente sección solo presentaremos los resultados pertinentes a la caracterización de la composición estructural de los viroplasmatas de rotavirus. Parte de productos resultantes de los estudios con los centros de replicación viral se pueden consultar en (Hidalgo et al., 2016).

Los resultados obtenidos por microscopia de súper resolución de viroplasmatas de rotavirus permitieron determinar la distribución de las proteínas virales VP4, VP6 y VP7, su ubicación relativa respecto a las proteínas virales no estructurales NSP2, NSP4 y NSP5, que participan en la formación de los viroplasmatas de rotavirus. De manera general, podemos decir que los viroplasmatas tienen una forma parecida a una esfera, aunque con desviaciones claras de este cuerpo geométrico (**Fig. 1**).

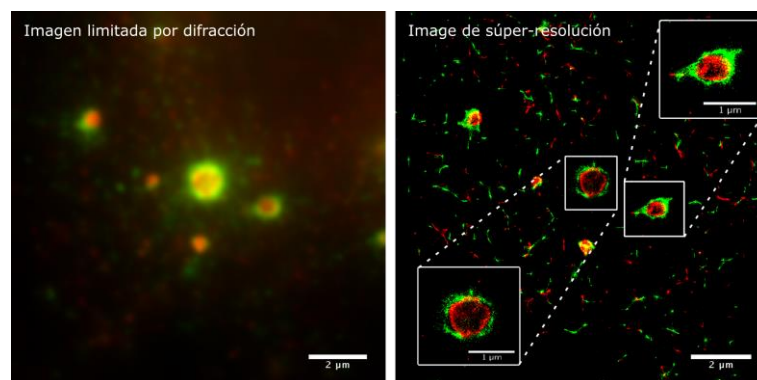


Figura 1. Composición estructural de los viroplasmatas de rotavirus. a) Imagen limitada por difracción de viroplasmatas de rotavirus en los cuales se observa a la proteína NSP2 que forma parte del núcleo del viroplasma (canal rojo) y la proteína VP6 (canal verde), componente del virión de rotavirus. **b)** Imágenes de súper-resolución obtenidas con el método 3B a partir de las imágenes limitadas por difracción.

Para simplificar el problema, propusimos un modelo de viroplasma equivalente a un corte sagital de una esfera. Bajo la hipótesis de que las distintas proteínas que conforman el viroplasma se disponen en niveles de organización que asemejan a los tejidos de una cebolla, desarrollamos un algoritmo que nos permitió cuantificar automáticamente la distribución relativa de las proteínas no estructurales (NSP4, y NSP5), y las proteínas estructurales (VP5, VP4, y VP7) a la proteína NSP2 (en experimentos pareados comparando siempre contra NSP2). El algoritmo, desarrollado a partir de (Garces et al., 2016), se centra en el ajuste de circunferencias concéntricas, donde se cuantifica el radio de los círculos correspondientes a la segmentación de cada distribución proteica a analizar, por ejemplo, NSP2 vs VP6 (**Fig. 2c - panel superior**). En este contexto, cuantificamos la distribución relativa de cada cuerpo, proteico o RNA de doble cadena, a NSP2 (**Fig. 2**).

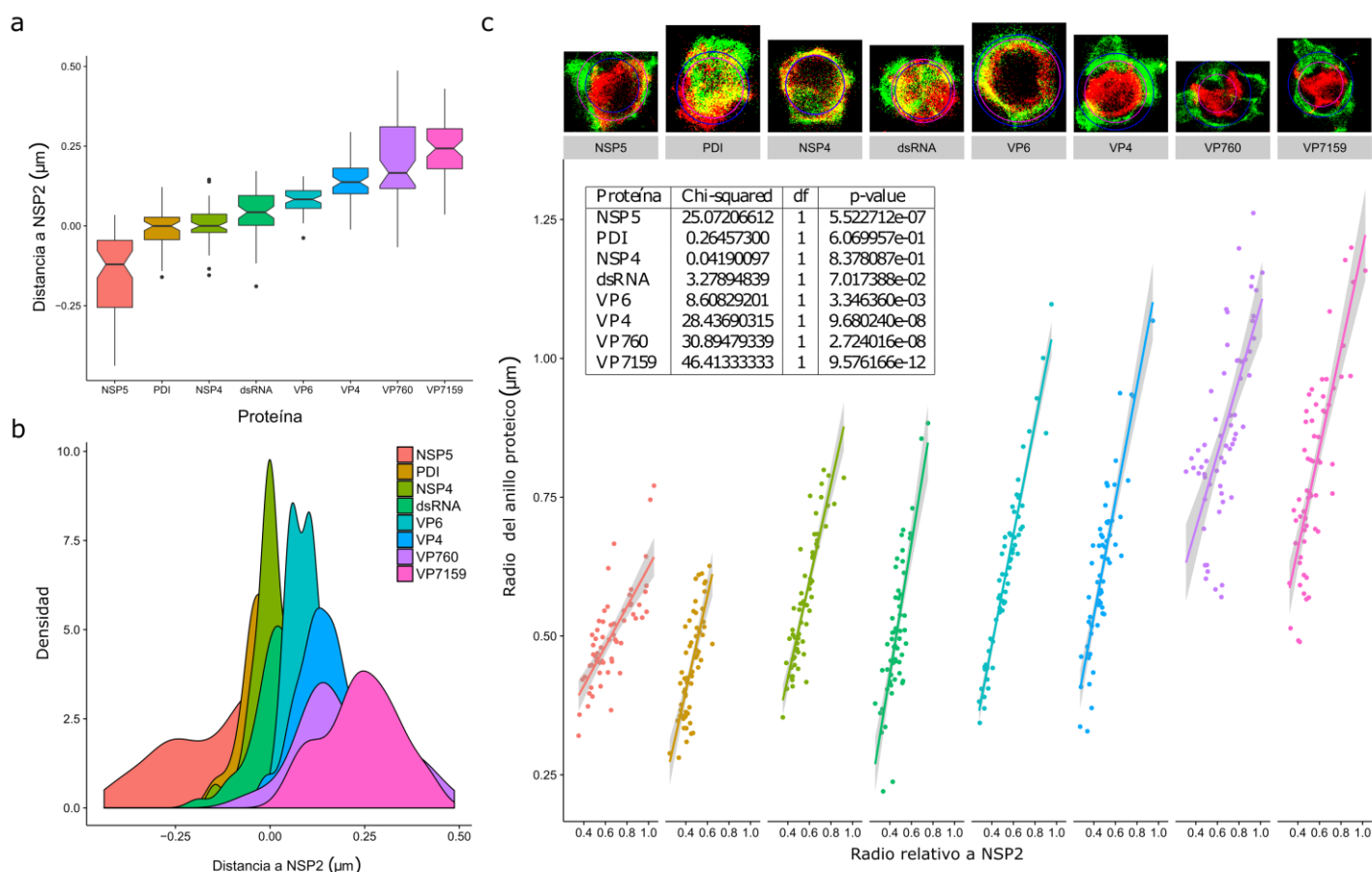


Figura 2. Análisis automatizado de la composición estructural de los viroplasmas de rotavirus. a) Gráficas de caja de la distancia que hay entre la proteína NSP2 y los demás cuerpos (proteínas, o RNA viral) contenidos dentro del viroplasma. Cada caja contiene el 50% de los eventos independientes, la línea interna muestra la mediana, y las barras de error delimitan el 95% de la distribución. Las bisagras muestran un intervalo de confianza alrededor la mediana $\pm 1.57 \times \text{IQRn}^{-1/2}$. Si las bisagras de dos cajas no traslapan hay evidencia de que las medias difieren (al 95 % de confianza) **b)** Gráfica de densidad que muestra las distribuciones de los datos presentados en (a). **c) Panel superior.** Los resultados presentados se obtuvieron mediante el desarrollo de un algoritmo 100% automatizado que detecta y cuantifica a los anillos que conforman a los viroplasmas, en experimentos pareados relativos al distribución espacial de NSP2. El resultado del ajuste de circunferencias para NSP2 se muestra en color rosa, mientras que el resultado del ajuste para cualquier otro cuerpo proteico (o RNA viral), se representa con un círculo azul. **c) Panel inferior.** Dependencia entre el radio de la distribución de NSP2 vs el radio de la distribución de los otros cuerpos proteicos. Note la relación lineal entre NSP2 y cualquier otra distribución (lo que apoya el modelo de las capas de cebolla). El inserto muestra los resultados pareados de la prueba Kruskal Wallis de cada cuerpo vs NSP2. Note que, el valor de p para NSP4 y DPI es inferior a 0.05, indicando que NSP4, NSP2 y DPI se distribuyen en la misma región dentro del viroplasma.

La distribución relativa entre las proteínas no estructurales indica que NSP5 se encuentra en el corazón del viroplasma, seguido un anillo formado por NSP2 y NSP4 (**Fig. 2**). Lo cual correlaciona con la función demostrada de NSP4 como proteína nucleadora de viroplasmas (Fabbretti et al., 1999).

La proteína VP6 se observa como un anillo externo a NSP2-NSP4 (**Fig. 2**); VP6 es la proteína que se ensambla sobre los corazones virales (formados por VP1, VP2 y VP3, además del genoma viral) formados en el viroplasma, dando lugar a partículas de doble capa denominadas DLPs (Estes *et al.*, 2013). Estas DLPs geman hacia el lumen del retículo endoplásmico (RE), por lo que la ubicación periférica de VP6 resulta lógica considerando que su ensamblado debe llevarse a cabo en los límites del viroplasma para permitir el contacto cercano con proteínas residentes del RE.

Actualmente, se postula que el viroplasma se encuentra rodeado por la membrana del RE, en donde residen NSP4 y VP7. Los resultados obtenidos nos han proporcionado información sobre la distancia y ubicación de estas proteínas virales con respecto al viroplasma. No obstante, consideramos necesario determinar la ubicación intracelular específica de NSP4 y VP7; con la finalidad de definir si las membranas en donde residen estas proteínas pertenecen al RE o a algún organelo diferente.

La ubicación de NSP4 en el interior del viroplasma, y en aparente interacción con NSP2, nos sugiere que NSP4 podría estar interaccionando con otras proteínas virales o celulares presentes en el viroplasma adicionales a VP6. Nuestros resultados indican que DPI, una enzima del RE, co-localiza con el anillo de NSP2-NSP4 (**Fig. 1 b, c**). NSP4 es una proteína residente del RE que se ha reportado interacciona con la proteína VP6 de las DLPs recién ensambladas permitiendo su gemación hacia el RE. Por lo anterior, la ubicación de NSP2, NSP4, NSP5 y DPI en el interior del viroplasma representa un hallazgo inesperado y sumamente interesante, ya que se hubiera esperado que NSP4, y el retículo endoplásmico *per-se*, se distribuyera de manera externa a VP4, al igual que VP7, ya que tanto VP7 como NSP4 son proteínas residentes del RE.

Interesantemente, el RNA de doble cadena de rotavirus se encuentra justo entre el anillo NSP2-NSP4-DPI y el anillo formado por VP6 (**Fig. 2**). Con lo anterior, proponemos que las DLPs se nuclean en las regiones mediales del viroplasma, en la interfase entre NSP2-NSP4-DPI y VP6.

Durante su gemación, las DLPs adquieren una tercera capa proteica conformada por la proteína VP7 y espículas formadas por la proteína VP4 (Estes *et al.*, 2013). A pesar de que ambas proteínas conforman en conjunto la tercera capa proteica, existen indicios de que VP4, una proteína citosólica, se ensambla sobre la DLP en un paso previo al ensamblado de VP7 (Gonzalez *et al.*, 2000). Nuestros resultados demuestran que, espacialmente, la proteína VP4 está ubicada entre la proteína VP6 al interior y la VP7 al exterior (**Fig. 1 b, c**), sustentando los reportes previos de que las DLPs entran en contacto primero con VP4 y posteriormente adquieren la proteína VP7.

En resumen, proponemos un modelo de viroplasma equivalente a un corte sagital de una esfera (**Fig. 3**), donde las proteínas nucleadoras NSP5, NSP2-NSP4 se encuentran al centro del viroplasma, estas últimas en interacción con componentes del retículo endoplásmico. En la región próxima al complejo NSP2-NSP4 se encuentra el RNA viral, que a su vez colinda con VP6. Finalmente, las capas externas del viroplasma están constituidas por VP6 seguida de VP4 y VP7.

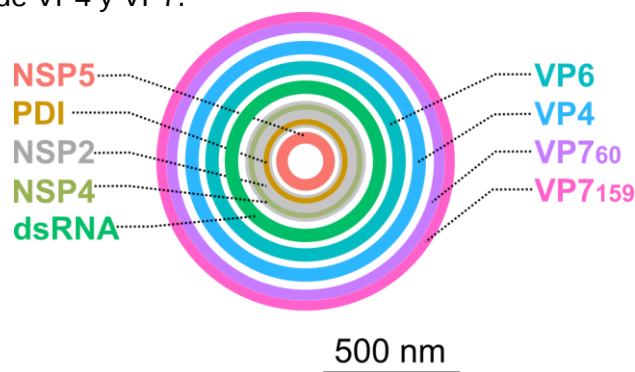


Figura 3. Modelo de la composición estructural de un viroplasma de rotavirus construido a partir del análisis de las imágenes de súper resolución. El modelo describe al viroplasma compuesto por una serie de anillos concéntricos de proteínas en donde, de adentro hacia afuera, el anillo más interno está formado por la proteína NSP5, seguido por un complejo NSP2-NSP4 que interacciona con componentes del retículo endoplásmico, ej. PDI, después el ARN viral (dsRNA), seguido de las proteínas virales VP6, VP4 y al final VP7.

Nuestros resultados fortalecen y enriquecen el conocimiento sobre la estructura de los viroplasmas de rotavirus. Sin embargo, se requiere de una mejor caracterización de esta estructura que hasta el momento permanece no elucidada. Consideramos necesario complementar este estudio mediante la determinación de la ubicación de otras proteínas virales que forman parte del viroplasma, como las proteínas virales VP1, VP2 y VP3 que

constituyen en conjunto el core del virión. Su localización en el viroplasma es de interés porque permitirá definir la zona del viroplasma en la cual se ensamblan los cores de los viriones.

Cálculos realizados (1 cuartilla)

Estadística Bayesiana en archivos multidimensionales de microscopía (x, y, tiempo, canal de fluorescencia).

Software utilizado

El análisis 3B genera un mapa de probabilidad de la localización de los fluoróforos a partir de un modelo ponderado resultante de todos los modelos posibles, donde cada modelo incluye diferentes números de fluoróforos, localización de fluoróforos, así como los estados dinámicos de los fluoróforos (Cox et al., 2012). La información almacenada en la secuencia de imágenes a analizar es comparada iterativamente con los modelos mencionados empleando inferencia Bayesiana (Cox et al., 2012). El resultado es una imagen de súper-resolución que permite resolver y cuantificar estructuras de decenas de nanómetros.

El algoritmo 3B calcula la distribución de probabilidad de las posiciones de los fluoróforos que mejor se aproxime a los datos experimentales, lo cual se logra empleando cadenas de Markov ocultas (HMM por sus siglas en inglés) y estadística Bayesiana (Cox et al., 2012). Esta metodología fue desarrollada por Susan Cox y Edward Rosten e implementada en el lenguaje de programación C++. El software está disponible en <http://www.coxphysics.com/3b/> bajo la licencia Pública General de GNU (GPL por sus siglas en inglés).

La implementación actual de éste método, que corre en un solo hilo, posee la desventaja de que el análisis es laborioso (horas a días). En el LNMA desarrollamos una plataforma de análisis en paralelo para el algoritmo de 3B empleando el mismo protocolo de paralelización conceptualizado para los servidores de la nube de Amazon EC3 (Hu et al., 2013). Paralelizamos el análisis 3B en subregiones de tamaño definido, considerando traslape, para una computadora personal, con un procesador IntelTM-Corei7-2700 CPU@3.4GHz, la unidad de súper-computo del IBt-UNAM (Hernández et al., 2016); y para la MIZTLI. En este protocolo, todas las subregiones a analizar se defienden previamente, posteriormente se ejecuta un script en consola que manda cada subregión a analizar. Cuando todos los núcleos están ocupados, la siguiente subregión permanece en la cola de espera hasta que se desocupe un núcleo. Lo anterior se repite hasta que todas las subregiones terminen de ser analizadas. La imagen de súper-resolución se obtiene finalmente de la concatenación de la información resultante.

Para la generación de las imágenes de súper-resolución modificamos la última versión liberada de 3B, threeB-1.1 (Hernández et al., 2016). Esto con la finalidad de establecer un tiempo definido de análisis ya que la versión threeB-1.1 no contempla un tiempo de paro definido, es decir el algoritmo corre hasta que el usuario decide detenerlo. Lo anterior se debe a que 3B hace un muestreo Montecarlo entre diferentes modelos, donde cada modelo corresponde a una cadena de Markov oculta (MCMC). En teoría éste proceso de muestreo puede tender al infinito, no obstante, los MCMC tienen una propiedad llamada “burn in” que tiene que ver con el tiempo que tarda el algoritmo en acercarse a la distribución de probabilidad real (en nuestro caso la estructura oculta en las imágenes de microscopía). Los autores recomiendan detener el análisis 3B a las 200 iteraciones debido a que en este punto el modelo ya convergió (http://www.coxphysics.com/3b/3B_releases/threeB-1.1/reference_manual/online-manual/index.html). Modificamos el código fuente de 3B para terminar los procesos a las 200 iteraciones.

Para compilar y ejecutar el software es necesario descargar e instalar los siguientes paquetes:

-TooN: <http://www.edwardrosten.com/cvd/toon.html>

-libcud: <http://www.edwardrosten.com/cvd/>

-gvars3: <http://www.edwardrosten.com/cvd/gvars3.html>

El software original fue probado para Linux y OSX, sin embargo, las modificaciones que hemos realizado solo han sido probadas en Linux (http://www.coxphysics.com/3b/3B_releases/threeB-1.1/reference_manual/online-manual/sComp.html).

Los ‘scripts’ de consola que desarrollamos para correr el algoritmo 3B en paralelo se pueden encontrar asociados a (Hernández et al., 2016), y se puede descargar de <https://bitbucket.org/hoha441/parallel-3b/src>.

Recursos utilizados

El análisis 3B demanda mucho tiempo de procesamiento para generar cada imagen de súper-resolución (SR), los resultados obtenidos dependieron en gran medida del uso de los recursos proporcionados por la Unidad de súper-cómputo de la UNAM.

El 3 de marzo del 2016 nos aprobaron 499968 horas CPU de las cuales empleamos solo un 28 % (140259.95 hCPU) para concluir con los objetivos planteados en la propuesta original. Lo anterior se debió a dos razones:

1. En el 2016 se publicó un nuevo algoritmo de súper-resolución que corre en paralelo, basado en arquitectura GPU (31,32).
2. En el 2016 adquirimos, con apoyo de la DGAPA al Dr. Christopher Wood (encargado del LNMA), un GPU Nvidia Tesla K40m, (arquitectura Kepler).

Migramos nuestros protocolos de análisis de microscopía de súper-resolución al GPU del LNMA. Como resultado, consumimos menos recursos de los que contemplamos en la propuesta inicial. Un resultado colateral fue el desarrollo de nuevos esquemas de análisis de microscopía de súper-resolución que fueron aprovechados por otros grupos de investigación del IBt-UNAM (Vega-Cabrera L-A, et al 2017).

El presente documento constituye el final de nuestras actividades en la MIZTLI.

Lista de colaboradores

El presente proyecto resulta de un esfuerzo colaborativo entre dos laboratorios grupos de investigación mexicanos expertos en virología de rotavirus y adenovirus dirigidos por:

- El Dr. Carlos Arias del Instituto de Biotecnología de la UNAM (IBt-UNAM)
- El Dr. Ramón González del Centro de Investigación en Dinámica celular de la Universidad Autónoma del Estado de Morelos (CIDIc-UAEMor).

Lista de artículos publicados

- Hidalgo P., Guerrero A., Wood CD., Valdes M., Dobner T., Gonzalez R. Morphological, biochemical and functional study of viral replication compartments isolated from adenovirus-infected cells. J Virol. 2016; 90, 3411-3427.
- Vega-Cabrera L.A, Guerrero A., Rodríguez-Mejía J.L., Tabche M.L., Wood C.D., Gutiérrez-Rios R.M., Merino E., Pardo-López L. Analysis of Spo0M function in Bacillus subtilis. 2017; Plos One. en prensa. (*producto colateral*).

Formación de recursos humanos

Licenciatura

- **David Torres Hernández.** Centro de Investigación en Dinámica Celular. Instituto de Investigación en Ciencias Básicas y Aplicadas. Universidad Autónoma del estado de Morelos (CIDIc-IICByA-UAEM). Participó en la colecta y el análisis de las imágenes de súper-resolución de viroplasmas. En fase de escritura de tesis.

Maestría

- **L.C. Haydee Olinca Hernández Aviña.** Estudiante de Maestría. Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE). Participó en el desarrollo de protocolos de paralelización de algoritmos de análisis para la microscopía de súper-resolución.
- **IBQ. José Luis Martínez.** Estudiante de la Maestría en Ciencias Bioquímicas (UNAM). Instituto de Biotecnología. Universidad Nacional Autónoma de México (IBt-UNAM). Participó en la preparación de muestras biológicas (células infectadas con rotavirus) y colecta de datos (imágenes) para microscopía de súper-resolución.
- **L.C. Raúl López.** Estudiante de la Maestría en Ciencias Bioquímicas (UNAM). CIDIc-IICByA-UAEM. Participó en la preparación de muestras biológicas (células infectadas con adenovirus) y colecta de datos (imágenes) para microscopía de súper-resolución.

Doctorado

- **M.C. Paloma Hidalgo.** Estudiante de Doctorado en Ciencias Bioquímicas (UNAM). CIDIc-IICByA-UAEM. Participó en la preparación de muestras biológicas (células infectadas con adenovirus) y colecta de datos (imágenes) para microscopía de súper-resolución.
- **M.C. Yasel Garcés Suarez.** Estudiante de Doctorado en Ciencias Computacionales. IICByA – UAEMor. Participó en generación de las imágenes de súper-resolución de viroplasmas y centros de replicación viral. Manejó y monitoreó todos los cálculos realizados en la MIZTLI. Analizó todos los datos correspondientes al presente informe.

- **M. C. Luz Adriana Vega Cabrera.** Estudiante de Doctorado en Ciencias Bioquímicas (UNAM). CIDiC-IICByA-UAEM. Participó en la preparación de muestras biológicas (bacterias), colecta de datos (imágenes) para microscopía de súper-resolución, y análisis de los mismos (*producto colateral*).

Congresos nacionales e internacionales

- Ninguno.

Divulgación

- En el No. 4,787 de la gaceta de la UNAM (26 de mayo 2016) se hace mención a los resultados derivados del presente proyecto: http://www.dgcs.unam.mx/boletin/bdboletin/2016_228.html
- Los resultados obtenidos se mencionaron en distintos medios de difusión, electrónicos e impresos, nacionales e internacionales: <http://www.asmtusa.org/index.php/general-science-blog/item/5337-super-resolution-reflections-on-microscopy-in-the-new-year>
- Reportaje sobre las actividades de I&D del LNMA en materia de microscopía de súper-resolución <http://www.conacytprensa.mx/index.php/ciencia/mundo-vivo/10741-laboratorio-nacional-de-microscopia-avanzada>

Referencias

- Cox, S., E. Rosten, J. Monypenny, T. Jovanovic-Talisman, D.T. Burnette, J. Lippincott-Schwartz, G.E. Jones, and R. Heintzmann. 2012. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature methods*. 9:195-200.
- Fabbretti, E., I. Afrikanova, F. Vascotto, and O.R. Burrone. 1999. Two non-structural rotavirus proteins, NSP2 and NSP5, form viroplasm-like structures in vivo. *The Journal of general virology*. 80 (Pt 2):333-339.
- Garces, Y., A. Guerrero, P. Hidalgo, R.E. Lopez, C.D. Wood, R.A. Gonzalez, and J.M. Rendon-Mancha. 2016. Automatic detection and measurement of viral replication compartments by ellipse adjustment. *Scientific reports*. 6:36505.
- Gonzalez, R.A., R. Espinosa, P. Romero, S. Lopez, and C.F. Arias. 2000. Relative localization of viroplasmic and endoplasmic reticulum-resident rotavirus proteins in infected cells. *Archives of virology*. 145:1963-1973.
- Hernández, H.O., P. Hidalgo, W. C.D., R. González, and A. Guerrero. 2016. Parallelizing the Bayesian Analysis of Blinking and Bleaching for Superresolution Microscopy. *In High Performance Computer Applications*. Vol. 593. 356-366.
- Hidalgo, P., L. Anzures, A. Hernandez-Mendoza, A. Guerrero, C.D. Wood, M. Valdes, T. Dobner, and R.A. Gonzalez. 2016. Morphological, Biochemical, and Functional Study of Viral Replication Compartments Isolated from Adenovirus-Infected Cells. *J Virol*. 90:3411-3427.
- Hu, Y.S., X. Nan, P. Sengupta, J. Lippincott-Schwartz, and H. Cang. 2013. Accelerating 3B single-molecule super-resolution microscopy with cloud computing. *Nature methods*. 10:96-97.