

Jóvenes científicos

La Facultad de Ciencias de la Universidad de Málaga cobija a grandes promesas de la investigación. Con intención de darles visibilidad surge Jóvenes científicos: una nueva sección de entrevistas realizadas por estudiantes de este centro para acercar a los científicos en ciernes a los que pocos años por detrás les andan. Inquietas cuestiones en el lenguaje directo de los jóvenes es la forma más sencilla de enseñar que la ciencia puede ser tan bisoña como buena. Para empezar, Francisco José Villena, estudiante de grado de biología, entrevista a dos futuros doctores en una disciplina que le fascina: la virología molecular.

La dificultad para controlar los virus radica en la gran capacidad de mutar sus genomas durante la infección lo que les permite evolucionar muy rápido. En sus estudios de doctorado, Luis Díaz y Pedro Seoane han desarrollado *QuasiFlow*, un programa informático que permite estudiar las secuencias genómicas que componen las poblaciones víricas, llamadas *cuasiespecies*, con un detalle que hasta ahora no era posible. Con él pretenden averiguar cómo se genera la enorme variabilidad genética de los geminivirus, unos virus de plantas

cuyo genoma está formado por ADN monocatenario que causan muchas pérdidas económicas en la agricultura mediterránea. Resulta impactante que para estudiar a los seres más minúsculos sobre la Tierra necesitemos a *Picasso*, una de las mayores supercomputadoras de España. El trabajo de Luis Díaz Martínez está supervisado por los Profesores Titulares de Genética Ana Grande Pérez y Enrique Viguera, mientras que Pedro Seoane realiza su labor investigadora bajo la dirección del Profesor Titular de Bioquímica Gonzalo Claros.



Luis Díaz (izquierda) y Pedro Seoane (derecha) en las instalaciones del supercomputador Picasso.

Créditos de la foto

FJ: En primer lugar, hablemos de vuestra creación ¿qué hace *QuasiFlow* desde un punto de vista informático? ¿Y desde el biológico?

PS: En sí mismo es un flujo de trabajo informático, lo cual implica que encadena un montón de tareas entre

sí para sacar información de las muestras de virus que tengamos. La intención es caracterizar la población vírica que contenga una muestra para saber cómo está y como se comporta dentro de su hospedador.

LD: *QuasiFlow* realiza un análisis de hasta 23 variables

que permite algo que hasta ahora era imposible en el estudio de cuasiespecies, el poder realizar experimentos con réplicas biológicas y comparar entre muestras. El concepto que se tenía de una cuasiespecie viral era que presentaban tal heterogeneidad que no era capaz de reproducirse en laboratorio para unas mismas condiciones, ya que cada una evolucionaba de forma distinta debido a la alta variabilidad del sistema. Como podías comparar un número muy reducido de clones virales era imposible ver sus similitudes, sin embargo nuestra aplicación es capaz de «normalizar» la cuasiespecie; crea una media de variables que permite ver los patrones comunes entre cuasiespecies que hacen que se comporten de la misma forma bajo un mismo hospedador.

FJ: ¿Qué ventaja supone vuestra aplicación frente a otras herramientas de estudio de secuencias genómicas preexistentes?

LD: En biología es muy importante la reproducibilidad y en eso se basa la fortaleza de nuestro software, consigue aportar a la comunidad científica una herramienta (de código abierto) de estudios y análisis de cuasiespecies más fuertes y reproducibles.

PS: Su automatización es clave; permite por ejemplo tener un análisis profundo de 24 muestras en paralelo (con recursos del supercomputador disponible) listo en 5h, mientras que antes tendría que dedicar hasta 120 días (5 días por muestra). Además, al ser automatizado permite poder analizar muchas más variables de golpe para detectar y correlacionar señales interesantes del estudio. Se gana enormemente en tiempo y cantidad de resultados con sólo descargar las secuencias y arrancar el programa.

FJ: Suena a que se trata de un programa muy potente al decir que se usa en una supercomputadora como *Picasso* pero, ¿puedo ejecutar *QuasiFlow* en cualquier ordenador?

LD: No, *QuasiFlow* puede necesitar hasta 75 núcleos y una memoria RAM que llegue a picos de 600GB... Los ordenadores comerciales más potentes suelen tener 16 o 32GB a lo sumo. Es imprescindible contar con una supercomputadora, como *Picasso*, donde están instalado todo lo que necesita *QuasiFlow* para trabajar. La idea es que cualquier investigador que lo necesite acceda remotamente a él en un único lugar.

FJ: Vosotros os habéis centrado en el estudio de variabilidad en geminivirus de interés agrícola, ¿es extrapolable al campo de la biomedicina, con virus tan cambiantes como los de la gripe?

LD: Sin duda, de hecho hicimos pruebas con virus de la hepatitis y el SIDA. *QuasiFlow* se centra en estudio de la variabilidad en una población viral, independientemente de quién sea o cómo se llame el virus o cual sea su hospedador. Se centra en estudiar la variabilidad que hay entre un número de secuencias que tú le ofreces,

siempre y cuando el virus ya haya sido secuenciado con anterioridad y podamos usarla de referencia para compararla con las que nosotros tenemos. Por ejemplo no serviría para trabajar con secuencias virales extraídas de una muestra de tierra, porque no podríamos trabajar debido a que se desconocen los virus que puede contener dicha muestra y no podríamos acceder a las secuencias de referencia para comprobar las variantes que existen en la secuenciación.

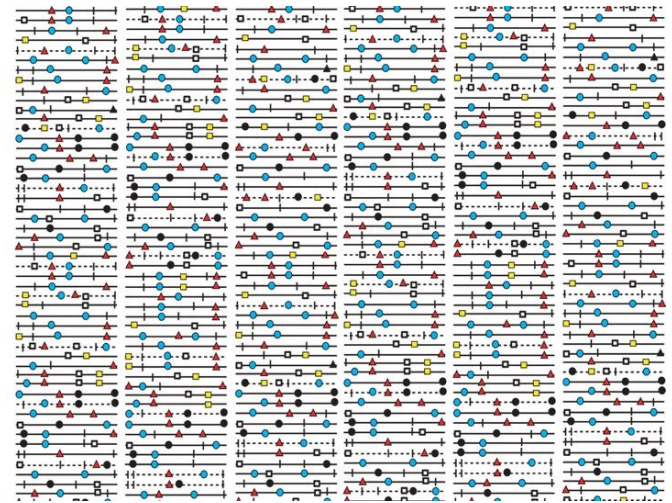


Figura 1. Si bien la imagen podría pasar por una obra neoplasticista del holandés Piet Mondrian, en realidad se trata de una representación gráfica del Dr. Esteban Domingo Solans, el virólogo español que obtuvo la primera evidencia experimental de *cuasiespecies virales*. Este término se acuñó para describir las distribuciones complejas de variantes genéticas que forman los virus ARN que están estrechamente relacionadas y que evolucionan para adaptarse a su hospedador. En la imagen se representa una fracción de la cuasiespecie: cada línea representa un genoma vírico y los símbolos sobre ellas marcan mutaciones. Las líneas discontinuas son secuencias incapaces de perpetuarse, debido a que superaron la tasa de mutación tolerada por la cuasiespecie. Por tanto, los virus de ARN no actúan como una única secuencia canónica, sino un conjunto de mutantes que evolucionan como unidad de selección. Un ejemplo muy recurrido por el autor de la imagen para comprender la magnitud de la variabilidad es el siguiente: «Son necesarias alrededor de 100 millones de imágenes como esta para representar la cantidad de mutantes del HCV (virus de la hepatitis C) que presenta un hígado humano infectado severamente por el virus. Pero eso no es todo; el número, tipo y distribución de las mutaciones es distinta en función del paciente». Sin duda, esta alta variabilidad hace que el estudio de estos virus y la obtención de secuencias consenso que faciliten su estudio para combatirlos sea un gran desafío. Para responder a dicho desafío se ha creado en la UMA *QuasiFlow*, una herramienta de análisis de datos de secuenciación masiva con firma malagueña. Foto cedida de «*Virus as Populations*» (E. Domingo 2016).

FJ: Quisiera cerrar la entrevista con una cuestión que personalmente me preocupa y estoy seguro que a más potenciales científicos de mi generación también: ¿Qué

panorama podéis describirme en cuanto a aspirar a la investigación en España en la delicada situación económica actual?

PS: Sinceramente, está complicado. Con los últimos recortes la financiación está muy mal, plantearte hacer un doctorado requiere algún tipo de beca además de hacer previamente un máster que tampoco es barato. En cuanto a la vía académica hay que tener las ideas claras y procurar un buen expediente para las becas. Por otra se necesita seguridad en uno mismo para tirar para adelante, es un camino duro que puede llegar a verse como un pozo sin fondo y creatividad para afrontar los problemas que la investigación plantea.

LD: Es muy importante además la «independencia» investigadora. La persona que se quiera dedicar a esto sepa que va a tener gente por encima suya a la que debe obediencia, pero nunca está de más, además, explorar por otros caminos siempre y cuando cumplas primero

con los que tus directores te marquen. El campo de la biología te permite afrontar los problemas de distintas formas, pero sin duda lo más importante es el esfuerzo para ser una persona curiosa, que se busque su hueco en un laboratorio y que su expediente le permita iniciarse en la investigación. Aquí las personas a las que les de un poco igual todo y no quieran hacer de esto su vida, ahora mismo como está la situación, no van a encontrar su lugar en la investigación. Es uno de los trabajos más gratificantes y cuando salen las cosas bien sientes como tu esfuerzo da resultados y merece la pena.

Parece que el desafío (Figura 1) que planteaban las cuasiespecies está más cerca de resolverse, ahora es a los potenciales biólogos a los que se nos presenta otro desafío; prosperar en nuestras pretensiones a la investigación y llegar a aportar a la comunidad científica un trabajo tan importante como el de los entrevistados.