

05/2008

## La bioinformática en el estudio de la diversidad genética



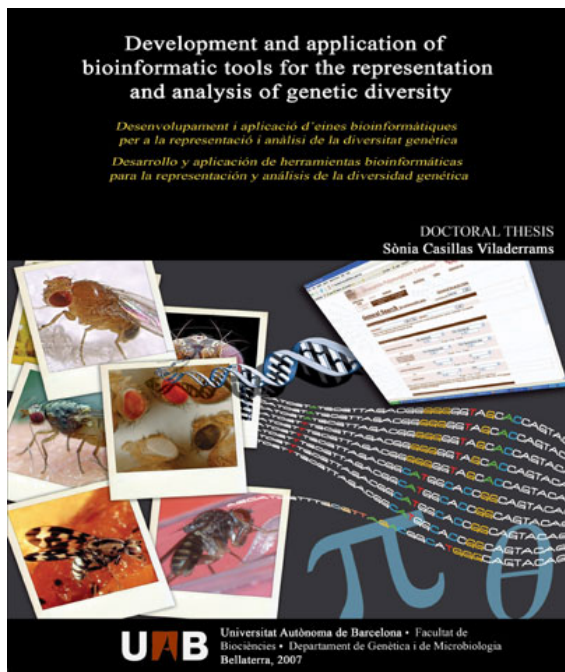
La variación genética es clave para entender la diversidad de seres vivos, así como las diferencias entre los individuos de una misma especie. Esta tesis estudia este fenómeno mediante el desarrollo de herramientas bioinformáticas para la extracción y análisis de secuencias de ADN, la generación de bases de datos de diversidad genética, y la comprobación de hipótesis a partir de secuencias de diferentes especies y regiones del genoma.

La variación genética es la materia prima de la evolución: son las pequeñas diferencias entre los individuos de una misma población las que, con el tiempo, generarán toda la diversidad de seres vivos existente en la Tierra. La descripción y explicación de las fuerzas que controlan la variación genética dentro y entre poblaciones es el tema central de la genética de poblaciones. La obtención de un número explosivo de secuencias de DNA en diferentes genes y especies ha cambiado radicalmente las perspectivas de este campo de trabajo, transformándolo desde una ciencia empírica insuficiente hacia un esfuerzo interdisciplinario de gran alcance, donde las estructuras de generación de nuevas secuencias a gran escala se integran con herramientas bioinformáticas para la extracción y gestión de datos, juntamente con avanzados modelos teóricos y estadísticos para su interpretación.

Esta tesis es un proyecto de bioinformática y genética de poblaciones completo, cuyo objetivo principal es el estudio de la diversidad genética en las poblaciones. Se ha llevado a cabo en tres pasos secuenciales: (i) el desarrollo de herramientas para la extracción, el procesamiento, la filtración y control de calidad de secuencias de DNA, (ii) la generación de bases de datos de conocimiento a partir de los datos obtenidos en la primera parte, y (iii) la comprobación de hipótesis que requieren de datos de varias especies y regiones del genoma.

En la primera parte de la tesis desarrollamos PDA (Pipeline Diversity Analysis, <http://pda.uab.cat>), una aplicación web de código abierto que permite la exploración de la variación genética a partir de los millones de secuencias almacenadas en las principales bases de datos públicas. Todos los datos extraídos y analizados en la primera parte de la tesis se utilizaron en la segunda para crear un recurso vía web de la variación genética en el género *Drosophila* (DPDB, *Drosophila* Polymorphism Database, <http://dpdb.uab.cat>). Finalmente, la tesis incluye dos estudios que utilizan los módulos de extracción y análisis desarrollados en la primera parte. En el primero de ellos estudiamos patrones de variación genética para inferir la acción de la selección natural en secuencias de DNA que no codifican ningún gen, pero que en cambio están conservadas entre diferentes individuos de una misma o de diferentes especies. Los resultados muestran que estas secuencias conservadas no codificadoras se mantienen gracias a la selección purificadora *Drosophila*. El segundo estudio se centra en la evolución de los genes Hox, una clase de genes involucrados en la especificación de las regiones a lo largo del eje anteroposterior del cuerpo y que son por tanto esenciales en el desarrollo. Sorprendentemente, los resultados muestran que los genes Hox evolucionan más rápidamente que otros genes que no son esenciales durante el desarrollo o que presentan patrones de expresión menos complejos.

En resumen, los trabajos presentados en esta tesis completan un ciclo de proyecto bioinformático, incluyendo todos los pasos necesarios desde la extracción de datos hasta la generación de nuevo conocimiento científico. Es más, el resultado de cada paso es la semilla para múltiples posibles estudios en el siguiente paso, y por tanto esta tesis tiene muchas aplicaciones para la comunidad científica.



*Figura 2. Portada de la tesis doctoral presentada por Sònia Casillas Viladerrams.*

### **Sònia Casillas**

Universitat Autònoma de Barcelona

[sonia.casillas@uab.cat](mailto:sonia.casillas@uab.cat)

## **Referencias**

"Development and application of bioinformatic tools for the representation and analysis of genetic diversity". Tesis doctoral presentada por Sònia Casillas Viladerrams el 22 de febrero de 2008 en la Facultad de Biociencias de la UAB y dirigida por los doctores Antonio Barbadilla y Alfredo Ruiz.

[View low-bandwidth version](#)