Área temática de Bioinformática

DOCUMENTO PRELIMINAR - MAYO 2003

Objetivos

Establecer las premisas para la constitución de un entorno de e-ciencia en el área de Bioinformática. Este documento se circulará entre los interesados para su comentario y refinamiento posterior.

Resumen:

Introducción

Definición del área:

En el contexto de este documento, usaremos el término de *Bioinformática* para incluir todas las *aplicaciones de la tecnología de la información (TI) en las Ciencias de la Vida*. Por tanto incluimos la aplicación de la TI en otras disciplinas conocidas con denominaciones afines como son la Biocomputación o la Biología Teórica cuya delimitación estricta no es objeto de este documento.

La integración de diversas aplicaciones de TI en Ciencias de la Vida bajo una denominación común está justificada por la creciente interrelación de las diferentes ramas de conocimiento y su progresiva confluencia hacia puntos comunes de encuentro.

Por motivos prácticos, se hace exclusión de la aplicación de la TI en las *Ciencias de la Salud* en esta sección. Si bien estas disciplinas están confluyendo e incrementando su sinergia recientemente y se espera una mayor imbricación en el futuro próximo, el problema de las Ciencias de la Salud es todavía suficientemente diferenciado como para garantizar un tratamiento separado, y por lo tanto es abordado en un Área Temática separada (Véase XXX).

Las Tecnologías de la Información en las Ciencias de la Vida:

El desarrollo de la Bioinformática ha ido estrechamente relacionado con el de las tecnologías experimentales. En este sentido, el desarrollo de la Biología Molecular ha supuesto un reto constante para las Tecnologías de la Información que han logrado a duras penas mantener un ritmo de crecimiento similar al de las Ciencias de la Vida en los últimos 20 años.

El desarrollo de las técnicas experimentales ha ido parejo a su popularidad en el ámbito científico y sobre todo a su repercusión e impacto social y económico, potenciado por hitos públicos de gran repercusión popular (como la secuenciación del genoma humano, las vacas locas y otros similares).

Ha sido precisamente uno de estos avances, la secuenciación del Genoma Humano, el detonante de la explosión de tecnologías experimentales conocida generalmente como *"Era Post-genómica"* (o "revolución de las *-ómicas"). En este contexto, se está abriendo un paradigma de trabajo nuevo que pretende abordar organismos enteros en lugar de genes o productos aislados:

las tecnologías de genómica, proteómica, transcriptómica, etc... permiten ahora realizar un abordaje holístico del estudio de los procesos que codifican la vida. Este cambio de paradigma supone un nivel de complejidad adicional al basarse en una integración masiva de enormes cantidades de datos sin parangón hasta la fecha.

Este salto tiene dos vertientes, una social y otra práctica. La primera y más seria es la toma de *conciencia social* sobre la factibilidad de abordar experimentalmente una serie de problemas hasta ahora inasequibles, con la consecuente presión sobre gobiernos y científicos para el desarrollo de estas tecnologías. La segunda y más grave es de índole práctica y significa un salto cuántico de varios ordenes de magnitud en el tamaño de los datos recopilados y otro aún mayor en el tratamiento de los mismos para alcanzar el grado de comprensión requerido para integrar datos de diversas fuentes en un enfoque sistémico.

El reto actual de todas las áreas de Bioinformática es responder a este cambio en la demanda de las necesidades analíticas en todas las áreas de las Ciencias de la Vida. Para ello no es posible continuar con los abordajes "clásicos" cuya capacidad analítica podía ir pareja al crecimiento exponencial de la tecnología.

El cambio está siendo progresivo, lo que está conduciendo a la mayoría de los grupos de Bioinformática a buscar soluciones a medida que los métodos en uso se iban quedando cortos. La primera fase de contención ha consistido en la instauración de sistemas de clusters y adaptación de algoritmos y procedimientos para su ejecución paralela a pequeña escala. Estas soluciones, si bien están permitiendo desarrollar y probar nuevos algoritmos y soluciones, resultan insuficientes para el tratamiento de datos experimentales masivos.

Las necesidades básicas se pueden expresar como una mayor demanda de capacidad de almacenamiento y tratamiento de información, y un crecimiento desmesurado de la capacidad de cálculo.

Los costes asociados al tratamiento de información derivada de las nuevas técnicas experimentales superan con creces la capacidad de prácticamente cualquier grupo o entidad aislada, aún recurriendo a soluciones paralelas de bajo coste como las granjas de PCs. En estas condiciones, la única solución viable consiste en la compartición de recursos entre grupos de forma solidaria hasta reunir recursos suficientes para abordar los problemas experimentales.

Los primeros resultados de estos esfuerzos están poniendo de manifiesto un problema adicional: la comprensión de las ingentes cantidades de información disponibles empieza a precisar de nuevos sistemas de *minería de datos* que permitan desarrollar métodos de inferencia sensibles al contexto capaces de extraer información relevante integrando fuentes *diversas y dispersas* de forma automatizada.

En otras palabras, *la forma más efectiva en coste y recursos de resolver los problemas actuales de las Ciencias de la Vida y por extensión responder a las crecientes demandas sociales sobre ellas es recurrir a sistemas masivamente distribuidos de tecnología Grid*.

e-Ciencia de la Vida

La aplicación de sistemas de e-Ciencia en Bioinformática no se justifica solamente por el crecimiento de las demandas computacionales. Además, precisamente por sus características, los problemas tratados proyectan perfectamente sobre una estructura computacional distribuida. Estas características son fundamentalmente dos:

- *Computación de grano muy grueso*: la mayoría de los problemas son divisibles en grandes subprocesos de intensa demanda computacional y ejecutables de forma independiente y en paralelo
- *Experimentos de muy alto rendimiento*: las grandes colecciones de datos de características similares generadas por las "*-ómicas" demandan un tratamiento uniforme (secuenciación en masa, análisis de estructuras masivo, etc..) y por tanto son tratables como una enorme colección de problemas independientes.

A esto debe unirse la enorme y creciente popularidad de estas técnicas experimentales reflejada en un gran número de laboratorios que aplican estas técnicas y son generadores independientes de grandes colecciones de datos. Esta situación aboga por una *organización descentralizada del almacenamiento, consulta y tratamiento de datos* apoyando con más fuerza la idoneidad de las tecnologías de Grid.

Necesidades básicas

Si bien hay una serie de problemas (procesamiento repetitivo de problemas masivos) que pueden empezar a beneficiarse inmediatamente de un entorno masivamente distribuido, aún no estamos en condiciones de explotar esta tecnología en toda su capacidad, precisamente por la novedad del problema experimental y consecuente escasez de abordajes informáticos.

Es de destacar que *ya existen grupos trabajando con tecnologías Grid*, especialmente en colaboración con otros grupos extranjeros y usando infraestructura foránea, pero sería irreal hacer depender las necesidades analíticas de la comunidad española de Ciencias de la Vida de Grids pertenecientes a otros países.

La situación de la mayoría de los grupos de Bioinformática en general puede calificarse como intermedia: han dado el salto a la *computación distribuida en clusters y granjas* de estaciones de trabajo, pero aunque en principio pasar los problemas a un entorno de mayor amplitud como es una Grid podría ser natural, no es posible afirmar que vaya a resultar trivial hasta que se aborde, y esto solo puede hacerse creando una infraestructura previa.

Finalmente, los usuarios, la principal fuerza de tracción sobre la Bioinformática plantean de por sí una problemática adicional: por un lado, pueden ya empezar a usar de forma inconexa algunos de los servicios desarrollados de forma distribuida sobre clusters, y en escasa medida sobre Grid. Además de necesitar nuevas herramientas y la adaptación de muchas más de las existentes, también precisan *puntos de entrada* que proporcionen un ambiente de acceso unificado a los diversos servicios disponibles, con *interfaces de trabajo intuitivos y versátiles*.

Participación en el 6º Programa Marco y colaboración internacional

La Unión Europea ha definido las tecnologías Grid como áreas prioritarias para su desarrollo en

el 6º Programa Marco. Más aún, tras el análisis de las expresiones de interés, ha dado especial relevancia a las iniciativas Grid en Ciencias de la Vida y la Salud, incluyéndolas en varias áreas del Programa, en ocasiones con precedencia sobre iniciativas Grid de más amplio interés. Esta decisión resulta llamativa cuando se considera la escasa representación de EoI en Ciencias de la Vida, dos en total, una de ellas remitida desde España en representación de la Red Europea de Biología Molecular (EMBnet).

Ya hemos mencionado la presencia de diversos grupos españoles de Bioinformática en iniciativas Grid extranjeras, tanto americanas como europeas. La participación de los mismos en éstas refleja la madurez de los grupos y la existencia y aprovechamiento de colaboraciones con grupos e instituciones pioneros de otros países, pero sobre todo refleja la existencia inmediata de necesidades de computación que no pueden ser satisfechas efectivamente en nuestro país.

Las fuertes relaciones existentes entre grupos españoles y extranjeros sitúan a los grupos de Bioinformática españoles en una situación bastante apropiada para la participación en iniciativas similares del 6º Programa Marco. De hecho, varios grupos están participando ya activamente en la preparación y presentación de iniciativas Grid para las primeras llamadas (como las propuestas EGEE y HealthGrid).

En otro orden, se están estableciendo las bases para constituir iniciativas Grid que incluyan a grupos de Bioinformática de países del entorno iberoamericano, bien dentro de colaboraciones auspiciadas por la UE como en otros marcos de colaboración internacional.

Por supuesto, es recomendable mantener y potenciar estas colaboraciones, fortaleciendo los vínculos que nos unen tanto a la UE, como EEUU, Iberoamérica y otros países, lo que conlleva una necesidad adicional: la de asegurar la compatibilidad e integración de cualquier iniciativa de e-Ciencia en España con sus equivalentes en otros países.

Un último factor a considerar es la razonable salud de la comunidad bioinformática española, favorecida por las actividades de la Red Temática Nacional de Bioinformática, en especial en relación con otros países del entorno, que la sitúa en un lugar de especial relevancia e influencia internacional; una situación temporalmente ventajosa que no sería aconsejable desaprovechar.

Áreas de trabajo y aplicación de una iniciativa de e-Ciencia de la Vida

Como ya se ha mencionado, nuestro objetivo no es realizar un ejercicio intelectual en bioinformática distribuida, sino responder a las necesidades de la ciencia experimental. En este sentido podemos replantear el problema en base a las distintas disciplinas y sus necesidades específicas:

- *Biología teórica*: modelización de sistemas biológicos complejos.
- *Bioinformática "tradicional"*: análisis de secuencias, y su extensión a la genómica.
- *Biocomputación*: análisis de estructuras, y su estensión al análisis de alto rendimiento.
- *Bioinformática "moderna"*: tecnologías de arrays y chips.
- *Tratamiento de información*: gestión de la información generada en el laboratorio: LIMS y bases de datos.
- *Provisión de servicios*: consolidación de servicios Grid de cara al usuario.

A éstas hay que añadir necesidades específicas de la Bioinformática en sí misma: a pesar de la familiaridad con tecnologías distribuidas en cluster, el entorno de Grid es suficientemente diferente como para precisar un apoyo de adaptación. Adicionalmente, el middleware de Grid aún no soporta de forma suficientemente transparente entornos de desarrollo avanzados como los precisos para poder asegurar un desarrollo eficiente en Ciencias de la Vida. En resumen, esto supone unas necesidades adicionales de:

- *Desarrollo de middleware y herramientas de desarrollo en Grid*
- *Formación y adaptación de desarrolladores a los nuevos entornos Grid*
- *Soporte en la instalación y mantenimiento de las facilidades locales*

Es de destacar que todas las áreas de aplicación descrita están representadas entre grupos españoles de Bioinformática de reconocido prestigio, así como que los intereses de los grupos a menudo se extienden sobre varias áreas simultáneamente. En la exposición siguiente mostramos las líneas de trabajo e interés de diversos grupos siguiendo de forma laxa la enumeración previa:

Rafael La Hoz -- Dpto. Biología Teórica, UCM

El Dpto. de Biología Teórica viene desarrollando desde hace tiempo modelos computacionales de estructuras macromoleculares y celulares complejas mediante autómatas finitos. La nueva disponibilidad de ingentes cantidades de datos a nivel de genomas o proteomas enteros y resolución estructural de alto rendimiento requiere para su comprensión la elaboración de modelos de simulación mucho más complejos y con un número muchísimo mayor de componentes.

El interés principal del grupo es la elaboración de experimentos de simulación *in silico* del comportamiento de estructuras celulares complejas. Por su alto número de componentes este tipo de experimentos de modelización son susceptibles de ser mejorados mediante técnicas eficientes de paralelización.

La disponibilidad de un entorno de computación distribuída proporcionaría mayores posibilidades de cálculo para los experimentos de simulación, permitiendo la realización de modelos con un número mucho mayor de componentes moleculares y por consiguiente mucho más completos, incrementando la comprensión de los procesos biológicos a escala celular.

Grupo de Julio Rozas, U. Barcelona

Este grupo trabaja principalmente en Evolución y Genética de Poblaciones Molecular. La evolución fué una de las primeras ciencias en usar tecnologías de la información, reflejo de su relevancia en la misma. El grupo trabaja en el desarrollo de aplicaciones para su uso aplicado y dispone ya de varios proyectos que requieren cómputo pesado:

1) Análisis de patrones evolutivos en genomas completos.

2) Estimación de parámetros poblacionales mediante simulaciones de ordenador basadas en métodos de Montecarlo.

El interés por tanto gira en torno a la adaptación de técnicas de análisis clásico a las nuevas tecnologías de genómica y la construcción de modelos avanzados para la comprensión de la dinámica de poblaciones.

Instituto Cavanilles de Biodiversidad y Biologia Evolutiva, Dept. de Genetica / Serv.

Bioinformatica. Universitat de Valencia

Este grupo ha empezado a trabajar en entornos distribuidos con el sistema de InnerGrid, aplicado a temas de reconstruccion filogenetica y genomica comparativa.
Aunque el proyecto se ha iniciado hace relativamente poco tiempo, cuenta con un estudiante de doctorado que va a trabajar especialmente con ello.

Laboratorio de Bioinformática. Centro de Astrobiología (CSIC - INTA).
Las líneas de trabajo de este grupo se centran, por un lado, en el estudio de la estabilidad termodinámica de proteínas y su relación con evolución molecular y modelos de plegamiento; y, por otro, en el análisis de interacciones entre proteínas y proteína-ligando.

El grupo se beneficia de la interacción con el laboratorio de Computación Avanzada del CAB y el departamento de Informática del INTA lo que está permitiendo el desarrollo de dos iniciativas en entorno Grid:

.- Una colaboración en curso, entre el Laboratorio de Computación Avanzada y el Laboratorio de Bioinformática, para realizar computaciones masivas con el programa PROTFINDER, desarrollado en el centro, sobre predicción de estructuras y propiedades termodinámicas de proteínas mediante métodos de alineamiento estructural (threading). El programa ha sido adaptado para su ejecución en un entorno Grid usando la herramienta GridWay, desarrollada también en el centro, y ya se han realizado las primeras pruebas. Actualmente, se están realizando algunas mejoras en el método para luego aplicarlo, mediante tecnología Grid, a grandes bases de datos de secuencias de proteínas, o bien, a genomas completos.

.- Un proyecto conjunto del Laboratorio de Bioinformática del CAB, el departamento de Informática del INTA y la empresa GridSystems para el uso de un gran número de ordenadores del campus del INTA en un proyecto de análisis masivo de interacciones entre proteínas implicadas en procesos de división celular bacteriana. En la actualidad se están realizando pruebas piloto utilizando el software InnerGrid sobre plataformas Windows, Solaris y Linux simultáneamente y una versión para Grid del programa AutoDock.

Grupo de Modesto Orozco. Univ. De Barcelona y Parque Científico de Barcelona

El grupo posee una larga tradición en la utilización de herramientas de cálculo intensivo enmarcadas dentro de la química cuántica, la dinámica molecular o el docking, a las que más recientemente se han incorporado de herramientas de bioinformática estructural que incluyen la construcción y manipulación de bases de datos estructurales. La principal limitación actual de este tipo de técnicas no es tanto el fundamento teórico de las mismas sino la capacidad de cálculo masivo. No tiene sentido, en la actualidad, analizar un número limitado de situaciones, cuando se dispone de información estructural suficiente para ampliar el estudio a situaciones generales. Para ello, los sistemas paralelos y en particular la informatica distribuida (GRID) constituye una solución excelente.

La disponibilidad de tecnología Grid ofrece la posibilidad tanto de optimizar la utilización de los recursos de computación ya disponibles, como de extender el uso de

herramientas tradicionales a sistemas de tamaño mucho mas realista.

Xavier Messeguer -- CIRI, Barcelona; R. Guigó, Alfonso González -- IMIM, Grupo de informática biomédica

En el entorno del IMIM y el Parque Científico de Cataluña han confluído una serie de grupos con un interés común por el desarrollo de aplicaciones y servicios en Bioinformática y la colaboración mutua. Todos ellos son grupos de sólida reputación y larga experiencia en la distribución de cómputo, con clusters de computadores instalados y en producción desde hace tiempo.

Algunos de estros grupos, como es el caso del de Roderic Guigó, están ya estableciendo relaciones con otras iniciativas Grid nacionales en otros países para desarrollar colaboraciones como una extensión natural a la distribución de sus trabajos (como blast o geneid) en clusters.

Evidentemente los grupos de Biologia Estructural también tienen un interés genuino en aprovechar iniciativas de computación distribuída: las aplicaciones de modelización molecular se cuentan entre las primeras en haber sido paralelizadas, con una larga casuística y experiencia en este problema, y la disponibilidad de una infraestructura masivamente distribuída multiplica el tamaño de los problemas que son abordables y -por consiguiente- extiende seriamente nuestra capacidad de comprensión del funcionamiento de los seres vivos.

J. M. Carazo -- Unidad de Biocomputación, CNB, CSIC

La Unidad de Biocomputación es un grupo de reconocido prestigio internacional con una larga experiencia en computación paralela (imparte esta asignatura en la UAM) que mantiene estrechas colaboraciones con otros grupos de computación de altas pestaciones nacionales y extranjeros, en especial con iniciativas Grid tanto europeas como Norteamericanas.

El interés del grupo se centra en la comprensión de procesos biológicos que requieren la interacción de grandes números de componentes que experimentan cambios estructurales dinámicos. Para ello es preciso desarrollar técnicas eficientes de resolución de estructuras de grandes complejos macromoleculares en especial mediante microscopía electrónica de transmisión.

La Unidad de Biocomputacion trabaja en el desarrollo de algoritmos de reconstrucción tridimensional que presenten un comportamiento robusto frente al ruido promediando un gran número de imágenes, desarrollando métodos iterativos que constituyen una alternativa ventajosa a los métodos tradicionales, pero que tienen demandas computacionales considerablemente superiores tanto en cómputo como en la necesidad ejecutar el mismo programa cientos de veces para obtener validaciones estadísticas.

La metodología de reconstrucción 3D constituye una de las aplicaciones que más fácilmente puede beneficiarse de las posibilidades de las grids, puesto que el empleo de las grid hace posible considerar estudios vedados hasta la fecha.

Adicionalmente la Unidad de Biocomputación trabaja en el desarrollo de sistemas para el acceso integrado a datos biológicos con una aproximación de mediación semántica. La resolución de heterogeneidades semánticas es esencial para la integración de información cuando se manejan datos complejos y cuando los usuarios y aplicaciones provienen de distintas disciplinas.

Grupo de J. L. Oliver -- U. Granada

Este es un grupo interdisciplinar de biólogos y físicos que estudia el genoma desde la perspectiva de los sistemas complejos. El grupo posee una sólida reputación como desarrolladores de aplicaciones de segmentación genómica que permiten descomponer un genoma en regiones de composición homogénea (isocoras). La delimitación precisa de las fronteras entre isocoras es muy útil para la anotación de genomas, aumentando sensiblemente la eficiencia de los programas de predicción computacional de genes y otros elementos funcionales (islas CpG, promotores, elementos repetidos, frecuencia de splicing alternativo, etc). Además, el grupo está también interesado en las correlaciones y en la estructura a gran escala del genoma, con la vista puesta en el análisis de las interacciones entre las partes que componen este sistema complejo. En ambas vertientes, el grupo viene experimentando una creciente necesidad de recursos computacionales distribuidos: por un lado están procediendo a paralelizar antiguos algoritmos secuenciales con el objeto de poder atacar en un tiempo razonable el análisis y la comparación de los genomas completos actualmente disponibles. Por otro lado, los nuevos algoritmos que están diseñando son ya paralelos desde el principio, utilizando la escasa infraestructura de computación distribuida disponible actualmente. El grupo mantiene colaboraciones con otros grupos, principalmente el Laboratorio de Evolución Molecular de la Stazione Zoologica Anton Dohrn de Nápoles, dirigido por Giorgio Bernardi, descubridor de las isocoras, y el Dpto. de Física de la Universidad de Boston, dirigido por Eugene H. Stanley, descubridor de las correlaciones de largo alcance en el ADN. Asimismo, mantiene una colaboración muy estrecha desde hace años con Wentian Li, miembro del Instituto de Santa Fe, Nuevo México, para estudios de la complejidad. Finalmente, el grupo tiene varias de sus aplicaciones disponibles en línea a través de la web, habiéndose convertido su servidor en el núcleo de la recién constituida Red de Bioinformática de Andalucía, por lo que precisan fuertes recursos para poder ofrecer acceso a algoritmos distribuidos a un número creciente de usuarios.

Javier de las Rivas -- U. Salamanca, Centro de Investigación del Cáncer, CSIC

El interés del CIC en las tecnologías de procesamiento paralelo le ha llevado a desarrollar entornos de producción distribuídos sobre las máquinas del Centro, llegando a convertirse en un prototipo de pruebas de la tecnología de GridSystems. Las áreas de aplicación de mayor interés son:

-Manejo de Datos de resultados de Genómica, especialmente de datos derivados de tecnología Affymetrix.

-Cálculo computacional sobre datos de expresión de microarrays genómicos aplicando herramientas y algoritmos con el paquete estadístico público R.

-Calculo computacional bioinformático, especialmente implementación de algoritmos tipo BLAST, Psi-BLAST, FASTA, HMM en entornos paralelos.

-Calculo computacional con paquetes de manejo biomolecular 3D: software de visualizacion, software de docking, software de dinámica molecular.

Alfonso Valencia -- Grupo de diseño de proteínas, CNB, CSIC

El Grupo de Diseño de Proteínas es uno de los grupos más representativos del país, con estrechas relaciones con muchos otros grupos de trabajo dentro y fuera de España. Desde hace años trabaja sobre entornos paralelos con granjas linux, y más recientemente con clusters Linux y HP/Tru64 y con un sistema Paracel BlastMachine de 20 procesadores. El grupo colabora con otras iniciativas

Grid europeas (como UK-grid) para poder correr algunos de los servicios que ofrece públicamente.
Además de el desarrollo de aplicaciones paralelas, el interés del grupo en tecnologías Grid se basa en la posibilidad de ofrecer servicios avanzados que requieren computación pesada:

Un metaservidor de estructura tridimimensional, llamado Libellula.

Un sistema de cálculo de las interacciones entre dos genomas, llamado ECID.

Un sistema de agrupación de secuencias en familias de proteínas, llamado FunCUT.

Un sistema de bases de datos para el análisis de genomas completos, llamado ORFandDB.

Dos servicios de análisis de abstracts de Medline, para encontrar los artículos relacionados con los genes proporcionados de entrada. Estos servicios son Geisha y HCAD.

Algunos de estos servicios son susceptibles de ser directamente integrados en entornos Grid desde el primer momento, como Libellula y ECID, asociados a la granja de PC's, y FunCUT, asociado tanto a la BlastMachine como la granja de PC's.

Dpto. de Arquitectura de ordenadores, U. de Málaga

Este Departamento posée una larga tradición de colaboración con grupos de investigación en Ciencias de la Vida y Bioinformática, trabajando en la paralelización de algoritmos pesados de análisis de datos biológicos. El Dpto. ha realizado una fuerte inversión en el desarrollo de aproximaciones basadas en multiprocesadores y procesamiento paralelo en diversas áreas. Sus líneas de trabajo básicas son:

- Diseño de entornos paralelos de computación -- arquitecturas distribuídas (Grid)
- Desarrollo de componentes software -- middleware
- Desarrollo de aplicaciones y paralelización de algoritmos.

En consecuencia, el interés es múltiple, centrado principalmente en el desarrollo de middleware y entornos paralelos, pero también en el desarrollo de aplicaciones y, en tanto que usuarios y proveedores de acceso a las mismas, consumidores importantes de fuertes recursos computacionales.

Servicio EMBnet/CNB

El Servicio EMBnet/CNB funciona como un servicio nacional a toda la comunidad investigadora del país, constituyendo el nodo español de la Red Europea de Biología Molecular (EMBnet) desde hace más de una década. Desde su proyección internacional mantiene estrechas relaciones con otras iniciativas Grid europeas y juega un papel central en la promoción y coordinación de iniciativas Grid en Ciencias de la Vida.
Los intereses fundamentales del Servicio se centran en la provisión de servicios de cómputo científico al usuario final investigador, la asistencia en el uso de las herramientas, la formación y el apoyo al desarrollo de aplicaciones.
La demanda de herramientas con alta demanda de prestaciones por parte de los usuarios empieza a notarse ya, y a medida que las nuevas técnicas experimentales se popularicen y extiendan es de prever un cambio consecuente: por un lado un incremento en el número de usuarios, y por otro, la incorporación de nuevos métodos de análisis altamente exigentes desarrollados por grupos de Bioinformática como los expuestos en éste documento.
En este sentido, es de notar que las herramientas desarrolladas en grupos de investigación suelen

precisar una elaboración posterior para facilitar su uso e integrarlas con las demás aplicaciones que requiere el usuario. El uso de aplicaciones diseñadas para Grid requerirá un esfuerzo importante para elaborar portales de acceso uniformes e integrar herramientas dispares en un entorno cómodo al investigador.
Finalmente el servicio constituye un punto de referencia y apoyo para los desarrolladores de Bioinformática. En este sentido se ha solicitado en el proyecto EGEE a la UE la asignación de dos ingenieros para proporcionar un servicio de formación, asesoría, soporte y adaptación de aplicaciones a entornos Grid que se ofrecerá a la comunidad bioinformática a nivel europeo que sería deseable incrementar con especialistas dedicados a apoyar específicamente a la comunidad española.

Esta lista no pretende ser exhaustiva, otras instituciones (p. ej. el Instituto Cavanilles de Biodiversidad y Biologia Evolutiva Dept. de Genetica / Serv. Bioinformatica, Universitat de Valencia, el Institut de Biotecnologia i Biomedicina Vicent Villar Palasi, UAB, el Instituto de Investigaciones Biomédicas del CSIC, etc..) y grupos (p. ej el grupo de Guillermo Thode en la U. Málaga) han expresado su interés y apoyo por una iniciativa de e-Ciencia para Ciencias de la Vida, bien como una extensión natural de sus recursos en cluster existentes, bien como consumidores de recursos avanzados de computación a distintos niveles.

Transferencia de Tecnología, Visibilidad y Difusión de Proyectos

Muchos de los grupos mencionados han demostrado ya su capacidad para transferir los resultados de su trabajo a otras entidades y generar patentes y productos comerciales, e incluso para formar el germen de empresas de nueva creación.

El creciente interés de la Biotecnología depende en buena medida para su éxito de la existencia de herramientas analíticas capaces de procesar las ingentes cantidades de información que se están generando. En estos momentos existe una gran escasez de aplicaciones, por lo que los proyectos enumerados son en su mayoría de gran interés tanto académico como comercial si pueden ser llevados a buen término. Esta es una razón más para fundamentar la importancia de la instauración de una iniciativa Grid: su presencia permitirá el desarrollo de aplicaciones necesarias y competitivas y reducirá nuestra dependencia tecnológica de grupos y empresas extranjeros.

El interés por las herramientas que hay que desarrollar procede de su utilidad y conlleva una repercusión importante: las aplicaciones desarrolladas deberán ser soportadas y mantenidas para su uso en producción. Este mantenimiento al no ser un trabajo de investigación no puede ser financiado con cargo a los presupuestos de investigación, lo que tiene dos efectos notables:

Por un lado supone un incremento en la relevancia de los servicios de apoyo a la investigación (como es el caso de EMBnet/CNB, que actúa como punto de referencia para otros servicios) en el papel de formación del usuario final, apoyo y asesoría técnica.

Por otro lado supone la necesidad de encontrar una entidad que pueda garantizar el mantenimiento y mejora progresivos de las aplicaciones más relevantes y populares, un papel en

el que la labor empresarial es fundamental.

En resumen, para que las iniciativas de desarrollo en Ciencias de la Vida tengan éxito en el futuro próximo hacen falta herramientas que aún hay que desarrollar y para las cuales resulta necesario disponer de una infraestructura de e-Ciencia. Para que esta iniciativa tenga éxito es preciso garantizar la transferencia de conocimientos y tecnología en varias direcciones:

- transmisión de experiencia y conocimientos de desarrollo paralelo a los desarrolladores desde servicios de referencia y otros grupos de desarrollo
- transmisión de software desarrollado a servicios y empresas, junto con el conocimiento preciso para explotarlo y mantenerlo
- transmisión del conocimiento de explotación y asesoría al usuario final desde servicios y empresas de apoyo.

Para garantizar estos procesos de comunicación conviene aumentar los intercambios que ocurren naturalmente con la dotación de recursos par la realización de reuniones y congresos que permitan el libre intercambio de ideas e información entre grupos, servicios y empresas (como por ejemplo las reuniones anuales de las Redes Temáticas), actividades de diseminación (como la creación de un portal WWW de referencia) así como la realización de cursos de formación para desarrolladores y cursos en el uso de aplicaciones para usuarios, y sobre todo, actividades de coordinación general de las iniciativas a través de uno o más coordinadores.