## Content

<table>
<thead>
<tr>
<th>Category</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EDITORIAL</strong></td>
<td>Interface Special Issue</td>
<td>3</td>
</tr>
<tr>
<td><strong>INTERVIEW</strong></td>
<td>The three pioneers who connected two worlds</td>
<td>4</td>
</tr>
<tr>
<td><strong>NBIC</strong></td>
<td>Consortium</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Toolbox</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Bioinformatics Network</td>
<td>39</td>
</tr>
<tr>
<td><strong>NBIC</strong></td>
<td>Research Programme, BioRange</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Support Programme, BioAssist</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Education Programme, BioWise</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Dissemination &amp; Exploitation, NBICommons</td>
<td>32</td>
</tr>
<tr>
<td><strong>BIOINFORMATICS</strong></td>
<td>Life Sciences &amp; Health</td>
<td>11</td>
</tr>
<tr>
<td>IN THE FIELD OF</td>
<td>Horticulture</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Food &amp; Nutrition</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Industrial Biotechnology</td>
<td>35</td>
</tr>
<tr>
<td><strong>THESES/PORTraits</strong></td>
<td>John van Dam</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Umesh Nandal</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Anand Gaval</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Jules Kerssemakers</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>RSG network</td>
<td>38</td>
</tr>
<tr>
<td><strong>INTERNATIONAL VISION</strong></td>
<td>Ewan Birney (Cambridge, UK)</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Philip Bourne (University of California, San Diego)</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Burkhard Rost (TU Munich, Germany)</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Amos Bairoch (University of Geneva, Switzerland)</td>
<td>29</td>
</tr>
</tbody>
</table>
Bioinformatics is driving the convergence of biology and technology disciplines, skills and infrastructures and is key to current data-intensive life sciences R&D. The broad field is under accelerated international development as biological systems turn out to be more complex than previously realised, and data sources are growing rapidly due to the ongoing flow of new key technologies. Re-use of large and information-rich datasets at an international scale is an emerging challenge. In fact, for any data driven research programme in contemporary life sciences, bioinformatics has become indispensable.

More often than not, the capacity and expertise needed to tackle bioinformatics challenges exceed the scope of the individual data generating institutes. To avoid stagnation of discoveries and to seize the innovation potential of its vital life sciences R&D sectors, the Netherlands needs a strong bioinformatics knowledge infrastructure. It is our mission to create an internationally operating centre of excellence in bioinformatics research and education that supports life sciences R&D.

In this special issue of Interface you will discover what the NBIC partners have achieved so far. Read about our activities in research, support, education, dissemination, and about the examples of where this has led to exploitation of research output by entrepreneurial bioinformaticians. All topics are illustrated with facts and figures, example projects and comments from our collaborating biology partners. In addition, as a reference for you to judge how far we have come in realising our mission three pioneers in the Dutch field of bioinformatics look back on how NBIC started. We have invited some international colleagues to comment on the latest developments in bioinformatics within the scope of the data-intensive life sciences.

A list of theses combined with some portraits of our junior researchers has been included to celebrate their key role in NBIC. They have contributed greatly to the scientific productivity of the NBIC groups, and they have been key to the success of a broad range of partner’s biology projects. As a growing community of our first generation PhD students, many have found one another through the Dutch Regional Student Group of the International Society for Computational Biology, a highly active group in the international field. Well trained within the NBIC faculty, the group of young NBIC scientists constitute the bioinformaticians of our near future and will no doubt play an important role in future discovery research. Hopefully, they will also bring the lively atmosphere in Dutch bioinformatics with them to their jobs in academic and industrial research.

Ruben Kok
The three pioneers who connected two worlds

The biologists and computer scientists had chosen seats at different tables, and they talked in different languages. Jacob de Vlieg, one of the three pioneers of the Netherlands Bioinformatics Centre, remembers the beginning of Dutch bioinformatics vividly. “We brought two worlds together.”

The secret of the team has to be that they complement each other so well. Because, apart from being scientists, the three pioneers of the successful Netherlands Bioinformatics Centre (NBIC) almost seem to come from different planets. The first is Gert Vriend, a highly creative biochemist and for many the face of Dutch bioinformatics. His natural habitat is his laboratory, working with sleeves rolled-up on model building and structure determination. Then there is Bob Hertzberger, a high-energy physicist by education. The high point of his scientific career was at CERN in the early 1980s, when he worked with then future Nobel Prize winners Carlo Rubbia and Simon van der Meer on the discovery of W and Z particles. He has been retired since 2006, but his home is a busy place with scientists walking in and out for good advice. Jacob de Vlieg, the man with the suit, completes the trio of NBIC pioneers. He adds the business touch, having worked for over twenty years at Unilever and Organon Research. Today he is the busy CEO of the recently established eScience Centre in Amsterdam.

These three men stood at the cradle of NBIC, the Netherlands Bioinformatics Centre, the Dutch network of bioinformatics experts active in research, education and support. How did they become involved, when and why? What were their roles? What did they learn? What have they enjoyed, and what has disappointed? And how do they picture the future of bioinformatics in the Netherlands?

THE BEGINNING

In 1999, Gert Vriend returned to the Netherlands from Heidelberg’s EMBL to establish and lead the Centre for Molecular and Biomolecular Informatics (CMBI) in Nijmegen. Vriend reminisces: “One of my tasks was to set up a national Dutch bioinformatics centre. It was kind of spectacular because NBIC became bigger and bigger in terms of budget. But it also became increasingly troublesome because everyone wanted to get aboard and people began to pull strings. I saw the best and the worst of my scientific colleagues in those days.” Bob Hertzberger was asked to assist Vriend in calming the hornets’ nest. He believes that building an institute like NBIC above all requires a clever architect, because the trick is in structuring the projects in such a way that everyone is at their best place. Hertzberger observes: “People may say that I’m short-tempered, but when scientists share my goal, I can be a very patient listener and puzzle together all the wishes.” Jacob de Vlieg joined NBIC’s pioneering team as the chairman of the NW0 research programme BioMolecular Informatics, which was to be merged into the new research
centre. “The Dutch industry saw great potential in the emerging field of bioinformatics, and I was very eager to join the scientific advisory board. But I broke out in a cold sweat at our first meeting with the field. Biologists and computer scientists had chosen seats at different tables, and real communication was hard to achieve; they talked in different languages.”

THE WORK. According to De Vlieg it was the ‘wild days’ of bioinformatics; the sheep had still to be separated from the goats. “Some biologists wanted the computer scientists to build websites for them; they really had no clue what bioinformatics was about. Therefore, it was even more important to formulate our targets and goals precisely, not only those for the scientific projects, but also those for the educational programme and the computational services that NBIC would encompass.” For some months, Gert Vriend, Bob Hertzberger and others met every week in Bob’s study in Bussum to discuss NBIC’s structure and to compile NBIC’s applications for funding. Vriend recalls: “We had fierce, but good discussions. Writing the proposals took an awful lot of time but in the end we succeeded in shepherding them through all the procedures.” One notable hurdle to be taken was to convince the CPB, the Netherlands Bureau for Economic Policy Analysis, to assign budgets from the Dutch natural gas revenues (BSIK) to the new field of bioinformatics. Vriend elaborates: “The competition was incredibly diverse; we were, for example, competing with a plan to drill under the North Pole ice.” The commission was also very different from the scientific boards Vriend had dealt with before. “I was very happy that we had Apple Reuver of IBM in our delegation; he spoke their language.” Above all Hertzberger remembers a very interesting time. The field of biology was quite new to him but immediately caught his interest. When he was asked to become the first director of NBIC in 2001, he was easily convinced. He remembers that he sometimes needed to take a
GERT VRIEND is senior professor of Bioinformatics of Macromolecular Structures, and the director of the CMBI bioinformatics institute at the Nijmegen Centre for Molecular Sciences. Vriend studied biochemistry at Utrecht University, the Netherlands. In 1983, he received a PhD from Wageningen Agricultural University for his studies of the assembly process of plant viruses. He worked as a postdoctoral fellow at Purdue, Indiana, USA and at the University of Groningen; and as a researcher at the EMBL in Heidelberg, Germany (1989-1999). He initiated the WHAT IF project, a program for visualising proteins. In 1999, he started the Centre for Molecular and Biocomputational Informatics (CMBI) in Nijmegen. His main research interests include: model building by homology; structure determination and verification; specialised databases; and the application of computers in (wet) biomedical research.

BOB HERTZBERGER holds an MSc (1969) and a PhD (1975) in experimental physics from the University of Amsterdam. From 1969 to 1983 he was on the staff of the High Energy Physics Group, later the National Institute for Subatomic Physics, and worked at CERN in Geneva with the Nobel Prize winner Carlo Rubbia. In 1983, he became a professor of Computer Science. In 2006, he became an emeritus professor, but until 2009, he was director of the Dutch e-Science project Virtual Lab for e-Science (VL-e), the predecessor of the Netherlands e-science centre. He established NBIC and was scientific director (a.i.) between January 2004 and February 2006. He continued his involvement with NBIC as deputy scientific director until March 2010.

JACOB DEVLEIG is CEO/CSO of the new Netherlands eScience Centre in Amsterdam, a joint initiative of NWO and SURF to enable large-scale analysis of ‘big data’. De Vlieg studied biophysics in Groningen and wrote his PhD thesis on the development of computational methods for biostructure determination. He worked at EMBL, Heidelberg to develop structural bioinformatics techniques. De Vlieg held research and management positions at Unilever Research from 1990-2001, after which he joined NV Organon as head of Molecular Design and Informatics. In 2006, he was appointed as VP R&D IT and in 2008 as global head of Drug Design & Informatics Schering-Plough, later MSD. Since 2000, he has been a part-time professor of Computational Chemistry in Nijmegen. From 2001-2008, De Vlieg was the chair of the NWO Biomolecular Informatics programme and from 2003-2006 he was the chair of the Scientific Advisory Board of NBIC.

firm stand. “Being independent of the bioinformatics community was an advantage to me. But I did not hesitate to set things straight either. It provided clarity and it just worked.”

THE RESULT This year NBIC rounded off its first ten years of existence, which to a major extent was funded by the BSIK funds. The scientific review was full of praise. “NBIC has created better genomics through better bioinformatics,” Hertzberger affirms. The Netherlands succeeded in keeping pace with the top countries. His main concern today, however, is that creative minds are not getting enough freedom to excel. “Top scientists are almost always persistent, stubborn people, and some may even be called peculiar. Increasingly, scientific institutes are led by managers who are not scientists themselves, and who often collide with head-strong scientists. This implies a danger of choosing average over excellence.” Gert Vriend is quite proud of NBIC. He feels Dutch bioinformatics has matured enough to stand on its own feet, provided that NWO recognises it as an essential part of biology. Although he would not choose to repeat the efforts they made to establish the institute, he has gained advantageous insights. “Building up NBIC has provided me with a much broader view of the field. I have realised that there is abundant opportunity for cross-fertilisation, which has resulted in the creation of professorships at CMBI in the health and food area.” According to De Vlieg: “It may seem almost absurd, but I think the biggest accomplishment of NBIC is that we got biologists and computer scientists talking to each other. We have brought two worlds together. That has taken considerable time, perhaps too much time, but we succeeded.” Nevertheless, he adds a warning: “The groups are still very vulnerable. If a bioinformatics expert leaves, they have serious problems finding a good replacement.”

THE FUTURE Vriend keeps fighting and warning against bureaucracy in bioinformatics and Dutch science in general. “The Netherlands spends more money on organising science than on actually doing it, if you ask me. Currently, nearly all grants I receive are foreign. Applied bioinformatics is doing well, but it is really difficult to do fundamental bioinformatics.” Hertzberger is convinced that bioinformatics will disappear as a separate discipline. “Forty years ago CERN employed forty statisticians, now there are none. Not because statistics is not important anymore, but because statistics has become an integrated part of physics, and because there are very good statistical programs at hand. In biology, the same process will occur; that is only natural.” De Vlieg foresees a change in focus in bioinformatics as well. China cannot be beaten in data acquisition; therefore institutes such as NBIC should focus on high-quality data mining and analysis. De Vlieg adds: “And we should not be averse to picking up the best bioinformatics tools around just because we ourselves did not make them. The real battle to win is implementing a coherent and sustainable bioinformatics environment for large-scale data analysis to help solve biological questions best.”

ACKNOWLEDGEMENT All three pioneers emphasised that there are many people who contributed to building NBIC in the early days. These include (in no particular order): Jan Willem Tellegen, Louis Verkleij, Margriet Brandsma, Herman Berendsen & Charles Buys, Apple Reuver, and Peter Folstar.
NBIC Consortium
A continuously growing cooperation supported by formal long-term partnership agreements with universities, university medical centres, research institutes and industry.

The open structure creates the flexibility to include new partners and faculty members. The NBIC consortium is establishing a supra-institutional bioinformatics hub that builds on a network of collaborating bioinformatics experts in academia and industry.

The NBIC Faculty forms the core of the community-based organisation. NBIC currently has 27 consortium partners and over 50 faculty members.

NBIC timeline

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>Raising funds</td>
</tr>
<tr>
<td>2003</td>
<td>BSIK / NGI-I</td>
</tr>
<tr>
<td>2004</td>
<td>NGI-I</td>
</tr>
<tr>
<td>2005</td>
<td>Embedding NBIC</td>
</tr>
<tr>
<td>2006</td>
<td>Cross discipline community building</td>
</tr>
<tr>
<td>2007</td>
<td>NBIC Consortium</td>
</tr>
<tr>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
</tr>
</tbody>
</table>

Partner agreements per January 2011, Faculty members per October 2011
Conquering the limits of biology

Bioinformatics research establishes the breakthroughs needed in cutting edge biology, and NBIC’s BioRange programme has laid the foundations for this approach. Marcel Reinders, professor of bioinformatics at Delft University of Technology and scientific director of NBIC/BioRange, and Marc van Driel, coordinator of BioRange, talk about the history, the achievements and the future of bioinformatics research.

When BioRange was started in 2003 to stimulate the bioinformatics field, there basically was no field to speak of. At the time, just a few scattered groups were engaging in bioinformatics research. Giving an impulse to this small and immature field required a long-term perspective and a pro-active approach. Marcel Reinders explains: “NBIC had to build the field almost from scratch. An open call for proposals did not seem the right approach. Instead, for BioRange I and II, NBIC defined themes and actively approached groups to participate. The funds were distributed to those groups that already performed bioinformatics research and groups that had the potential to develop such activities.” In total, around 95 researchers have been funded by the BioRange programme. To outsiders, the position of bioinformatics research is sometime elusive. Is it focused on cutting edge computer science, on tool development, on contributing to biology? According to Reinders, it is all of the above. “There is always a link with biological research questions; within BioRange we do not perform blue skys research.” According to “We don’t do blue skys research” Marc van Driel, BioRange research is fuelled by truly advanced research in biology. “We operate where biology approaches the limits. BioRange is about bioinformatics research that enables biology to cross borders every time.”

MIXED OUTPUT The output of the BioRange programme reflects the clear choice to focus on cutting edge biology. Reinders: “The number of papers resulting from BioRange projects has exceeded our expectations. The Mid-Term Review revealed that the citation level of NBIC-associated publications is twice the international average.” He emphasises that not only is the quality of the papers high, but that they also clearly reflect the multidisciplinary nature of the field. “It is a really good mix of biology papers in leading science journals and publications in specialised bioinformatics journals. Clearly, progress in biology requires advanced bioinformatics research.” According to Reinders, the output in terms of tool development has not lived up to the expectations. Van Driel agrees: “It does not mean that all the tools resulting from research projects should be developed further, but there is certainly unused potential here.”

NEXT GENERATIONS For a research programme, the number of papers is a natural indicator of success. But both Reinders and Van Driel point out that they see the establishment of a vibrant research community as perhaps the most important achievement of BioRange. “Building a community and educating the next generations...
of bioinformaticians are very important elements of BioRange,” says van Driel. Most of the researchers funded by BioRange are PhD students – the future leaders in the field. Building a community is in part a natural result of ensuring large-scale research projects that require contributions from many researchers operating in different areas. Other instruments are the twice-yearly BioRange meetings, one of which is part of the NBIC Annual Conference where all BioRange researchers come together. A recent initiative are the ‘hot topics’ sessions, small-scale informal meetings that address a specific technology or development relevant to bioinformatics. Van Driel says: “These sessions are open to researchers outside BioRange as well, which again stimulates community building.”

“BioRange is about enabling biology to cross borders”

LEAVING THE WET-LAB  Having now established itself as invaluable to biology, what are the next peaks for bioinformaticians to climb? Bioinformaticians (and others) often joke that their ultimate goal is to make themselves redundant. To Reinders, this perceived redundancy is in fact the final removal of the artificial border between the two areas. “There is no way we can attempt to understand the complexity of biology without the use of computational methods. In biology, the wet-lab will gradually lose its importance in favour of computers. We are redefining the experiment. This no longer relates to your physical operations in a lab, but to the way you handle and process your data. In essence, we are moving towards a world where bioinformatics is not only part of biology, it is biology.”

When asked about his BioRange research project ‘Curated database for integrating a wide variety of genome-scale data’, postdoc Patrick Kemmeren has to search his memory. The project was finished in 2008, but he digs back even further. “Around 2001, the first large-scale datasets became available for protein-protein interactions, and we were looking for new ways to exploit these sets so we could extract more biological information.”

Kemmeren and his colleagues took the first step towards the now hot topic of data integration. They combined data on protein-protein interactions with gene expression data to determine a mark of confidence for the measured interactions. “This was a highly innovative approach. We succeeded in creating added value from the data.” Step by step, they expanded their approach to cover all types of different data, including 1,124 gene expression profiles; 54,949 protein-protein interactions; 1,195 phenotype conditions and subcellular localisation of almost all proteins. The integrative dataset was born. Working with the University of California San Francisco, Kemmeren and colleagues developed an algorithm to determine a ‘confidence number’ for each of the 401,820 protein-protein interactions found in two separate large-scale mass spectrometry datasets. Kemmeren: “This integrative data warehouse is still one of the most cited datasets related to stable protein-protein interactions. It provides a tool to rank your results and focus on the interactions that are most interesting for your particular hypothesis.”

“This was a highly innovative approach. We succeeded in creating added value from the data.”
Being allowed to do more than you thought at first; that is one of the intriguing results of the BioRange research project ‘Developing clinical predictors based on high-dimensional genomics data, pathway information and directed experimentation’. Jelle Goeman, then postdoc and now head of the Bioinformatics Expertise Centre at the LUMC unveils the mystery.

Analysis of gene expression data could be more straightforward by concentrating on groups of genes that exhibit a similar function, says Goeman. “If you include that information in the analysis, you gain statistical power because you reduce the so-called multiple testing problem. To ensure that the number of false positives is as low a possible, you have to perform a lot of tests. With each test, you lose statistical power. Working with groups of genes reduces the number of required tests, which also reduces the loss of statistical power.” This relationship was already known, but Goeman and postdoc Livio Finos were the first to exploit it in a more systematic way.

The main result of their work is that applying a sequential approach to multiple testing is the key to countering the drawbacks of testing. Goeman states: “You do not need to apply stricter rules with every testing step, but you can ignore uncertainties from previous steps without exceeding the margin of error. Using our findings, researchers can build multiple testing methods that use a priori information much more effectively. We showed that you actually are allowed to do lot more than previously thought.”

“You gain statistical power because you reduce the so-called multiple testing problem.”

Understanding the relationship between genotype and disease phenotype has become a mission for many biomedical researchers. Identifying the relevant genes for just one single disease is already a daunting challenge, but apparently researchers should open their minds to other diseases as well. In his thesis “Phenotype-guided disease investigation using bioinformatics,” bioinformatician Martin Oti showed how using existing knowledge on similar symptoms could seriously improve the search for new disease genes.

“Similar symptoms from apparently unrelated diseases can point to similar affected genes.”

“Let the symptoms speak”

“My research was fuelled by the Human Genetics department. They study rare diseases and were looking for an efficient way to select the most promising candidate genes,” Oti explains. When it comes to rare diseases, existing information may be sparse. “There is a host of genotype-phenotype data available in the literature, but we found that using this data does not always offer good predictions for rare diseases. Genes related to well-studied diseases are overrepresented and distort the results.”

Working with raw, unbiased data offered better predictions, Oti found. “In spite of all the ‘noise’ in these large datasets, we demonstrated that this data is very useful. An interaction with a known gene from a disease with similar symptoms increases the chance of identifying the right gene in a certain genomic region by a factor 10.” According to Oti, diseases are better viewed not as separate entities, but rather as collections of symptoms. “It is not simply black or white. Similar symptoms from apparently unrelated diseases can point to similar affected genes. Everything is connected.”

“Similar symptoms from apparently unrelated diseases can point to similar affected genes.”
Because of ageing, changes in lifestyle and the rise of chronic diseases, our demand for healthcare is changing. The accelerated pace of scientific and technological progress is key to anticipating the changing needs. One of the main drivers is genomics research, which provides increasing insight into the role of genetic factors in human health and diseases. High-throughput molecular analyses, including whole genome sequencing, allow the study of complete genetic maps and expression patterns of every gene. The understanding of genetic variation and the resulting changes in genes, proteins and metabolic pathways will yield biomarkers for the prediction of illness and drug response.

However, computational technologies are needed for processing, analysing and explaining the vast amount of data resulting from second generation sequencing, advanced proteomics techniques and complex metabolomics research. Advanced methods for data analysis, including biostatistics, machine learning, text mining and many more should be applied in order to increase our understanding of structural and functional genomics. Bioinformatics approaches enable unlocking the data in order to find answers to the essential biological and health questions of today.

**FAST GROWING FIELD OF BIOBANKING**

“Next generation sequencing and ‘omics’ technologies have turned the biobanking field – both for common and rare diseases – into one of the fastest growing areas of life sciences. To store, decipher and functionally elucidate the tremendous data flow, bioinformatics has moved to centre stage. Strengthening its infrastructure and human resources is critical for turning data into insight and yielding better diagnostics, therapy and prevention.”

*Gertjan van Ommen (LUMC)*

**MODELLING DISEASE DEVELOPMENTS**

“Large scale longitudinal tissue and clinical data collection and computational analysis of the molecular data derived from these samples are essential to model disease development under therapeutically relevant conditions. Without such models, the war on complex diseases such as cancer will remain an uphill battle with too many casualties.”

*Lodewyk Wessels (Netherlands Cancer Institute)*
PhD theses

1. Bayesian Markov random field analysis for integrated network-based protein function prediction
   Yiannis Kourmpetis
   Wageningen University, 4 October 2011
   Promotor: Prof. dr. E.J.F. ter Braak | Co-promotor: Dr. R. C.H.J. van Ham

2. Exploratory pathway analysis
   Thomas Kelder
   Maastricht University, 8 July 2011
   Promotor: Prof. dr. F.J. van Schooten | Co-promotor: Dr. C.T.A. Evelo

3. Analysis of metabolomics data
   Norman Droisma
   Leiden University, 10 May 2011
   Promotores: Prof. dr. J. van der Greef, Prof. dr. T. Hankemeier, Prof. dr. J.J. Meulman

4. Fish genomes: a powerful tool to uncover new functional elements in vertebrates
   Elia Stupka
   Leiden University, 11 May 2011
   Promotor: Prof. dr. J.N. Kok | Co-promotor: Dr. F.J. Verbeek

5. Genome scale prediction of protein subcellular location in bacteria, with focus on extracellular and surface-associated proteins
   Maomiao Zhou
   Radboud University Nijmegen, 4 January 2011
   Promotor: Dr. R. J. Siwan

6. Evolution of Ras-like GTPase signaling pathways
   John van Dem
   Utrecht University, 30 March 2011
   Promotor: Prof. dr. J.L. Bos | Co-promotor: Dr. B. Smal

7. Aspects of ontology visualization and integration
   Julia Dmitrieva
   Leiden University/LdACS, 14 September 2011
   Promotor: Prof. dr. J.N. Kok

8. Computational approaches for dissecting cancer pathways from insertional mutagenesis data
   Jaron de Ridder
   Delft University of Technology, 31 January 2011
   Promotor: Prof. dr. M.J.T. Reinders | Co-promotor: Dr. L.F.A. Wessels

9. Fluxes of Life - Bioinformatics for Metabolic Flux Quantification in Isotopic Non-Steady State
   Thomas Binsl
   VU University Amsterdam/IBIVU, 11 March 2011
   Promotor: Prof. dr. J. Herings | Co-promotor: Dr. J.H.S.M. van Beek

10. Work flows in life science
    Ingo Wassink
    Twente University of Technology, 8 January 2010
    Promotores: Prof. dr. A. Nijhuis, Prof. dr. G.C. van der Veer
    Co-promotor: Dr. P.E. van der Vet

11. Penalized canonical correlation analysis: unraveling the genetic background of complex diseases
    Sandra Wassenberg
    University of Amsterdam, Faculty of Medicine, 8 June 2010
    Promotor: Prof. dr. A.H. Zwinderman

12. Improving breast cancer outcome prediction by combining multiple data sources
    Martin van Vliet
    Delft University of Technology, 6 April 2010
    Promotor: Prof. dr. M.J.T. Reinders | Co-promotor: Dr. L.F.A. Wessels

13. Patterns that matter
    Mathijs van Leeuwen
    Utrecht University, 9 February 2010
    Promotor: Prof. dr. A.P.J.M. Sibbes

14. Vertical integration of high-throughput measurements to derive functional and regulatory interactions in S. Cerevisiae
    Rogier van Berlo
    Delft University of Technology, 13 September 2010
    Promotor: Prof. dr. M.J.T. Reinders | Co-promotor: Dr. L.F.A. Wessels

15. Proof of Concept: Concept-based Biomedical Information Retrieval
    Dolf Trieschnigg
    Twente University of Technology, 1 September 2010
    Promotores: Prof. dr. F. M. G. de Jong, Prof. dr. W. Kraaij
A half year ago he defended his thesis on the evolution of Ras signalling pathways. His attempts to unravel the evolution of these pathways, which regulate a wide variety of cellular process, were pretty productive. “I have been very lucky that the project I worked on ended in good, publishable results. Little was known in this field of research when I started. That was a big advantage. So almost every time I found something it was new. Of course I was surrounded by good coaches,” says John. Berend Snel of the Bioinformatics group at the Utrecht University was a major source of support and guided the way. Hans Bos of Molecular Cancer Research, UMC Utrecht gave feedback on the biological relevancy of the results.

John looks back with pleasure on his time spent doing his PhD. “Bioinformatics is a growing field where you can find your place. Very reassuring. At the end of my PhD I noticed that research groups that had run into problems because they did not know how to analyse their data were now using bioinformatics. And the discipline is still growing.” John has experienced the annual NBIC conferences to be a good way to get connected to the network of bioinformatics. “It is also the place to show more of yourself and your results. That is important for the future. People remember you better when they have seen your face,” he explains.

And it works. Before he finished his thesis he found his job: a postdoc position at Martijn Huynen’s Comparative Genomics group at the CMBI, Nijmegen. “When I saw the job description hanging on the bulletin board, I thought: this is me. This is what I want to do.” So John sent in his application and ended up the favourite. He now works on a completely new subject: the evolution of cilia and flagella, as part of the European SysCilia Project. Since the CMBI is on the same floor as the NBIC, he easily keeps in touch. “I continue to follow the NBIC conferences to keep meeting interesting people and to hear all about new developments.” His ambition is to become a group leader with a specific expertise in comparative genomics one day. But first, he will work on a good set of publications in well-established journals.

John van Dam, PhD: ‘Evolution of the Ras-like signalling pathways’; Thesis, 2011, Utrecht University; Promotor: Prof. J.L. Bos | Co-promotor: Dr. B. Snel
Demonstrating the new scientific mindset

Barend Mons, scientific director of NBIC/BioAssist, and Rob Hooft, coordinator of BioAssist, are complete opposites. Still, they share a mission: that of establishing a new scientific modus operandi based on community building. Data generation in itself is no longer the road to progress. Creating the right environment for extracting knowledge is the key to tackling the complexity of biology.

Describing BioAssist as NBIC’s technical support programme is cutting the corners a bit. The driving philosophy behind BioAssist aims for more.

Almost in unison, Mons and Hooft jump at the chance to pitch their five-word tagline: making other people’s data work. That requires a variety of technical solutions, but above all, it requires a new scientific mindset. Mons: “If we want to make sense of this biological complexity, we need new ways to collaborate, but the old way of doing science still prevails.” Hooft illustrates this: “Scientists normally operate like ‘I have a question, therefore I will measure something and that will give me my answer’. But answers can also be found in the huge amount of existing data, and sometimes a computer can even find answers to questions we did not know we had.”

“Computers can even find answers to questions we did not know we had”

TANDEM The overwhelming amount of data and the intricate way in which everything is connected is simply too much for individual scientists, according to Mons. “No one is clever enough to make a difference all on his own. You need to operate together to tackle these questions. What we need is respectful collaboration between scientists.” This implies acknowledging the gap between biologists on one side and computational-oriented scientists on the other. Mons explains: “That gap is real. It may diminish over time, but it will never disappear. The NBIC management therefore represents a mixture of these fields and we consciously try to navigate this gap.” The way they operate in tandem shows that a gap does not have to be a problem. Hooft clarifies: “Barend and I are complete opposites, but it works. He is the dreamer who plots the course.” Mons joins in: “And Rob is the one who makes sure we stay on course and that people deliver. He can discuss computational matters on a PhD level; I am the biologist of the two. I know nothing of computers. When my laptop crashes I call Rob.”

PLACE TO BE The new approach advocated by Mons and Hooft also requires a community of professionals that support and enable research. Not in the old way of an individual bioinformatician working within an individual research group, but as a community of bioinformaticians capable of taking science to the next level. Hooft: “In every biology group you will find a bioinformatician, who is generally much appreciated. But collaborating on bioinformatics issues has not been considered worthwhile. Group leaders are generally inclined to keep ‘their’ bioinformatician indoors and only let him or her work on problems directly relevant to the group.” Both Mons and Hooft feel that breaking through this attitude and creating a community of...
bioinformaticians, both researchers and scientific programmers that really work together, is their main achievement. Hooft says: “Make people work together on problems, not only on paper, but physically as well. That is crucial.” The monthly BioAssist meetings have become the epicentre of the Dutch bioinformatics field. Mons: “Bioinformaticians from all over the Netherlands come to our monthly meetings. And not because they have to, but because they really want to participate. These meetings are the place to be.” And the crowd keeps on expanding, says Hooft. “Apparently, we have set up something that offers added value to a lot of people in this field.”

**“No one is clever enough to make a difference all on his own”**

So far we haven’t touched upon the technical details of the BioAssist operation. What does the programme entail in practice? Hooft doesn’t need a lot of words: “We look at what is technically possible and make that usable and available to the field at large. It is all about employing existing possibilities. Actually, the core of BioAssist is that we don’t do anything new.” They both start to laugh. Mons: “In fact, doing nothing new is the most innovative aspect of BioAssist.”

**Facts & Figures**

**Support**

- **25** Scientific programmers
- **10** Postdoc positions
- **5** Dedicated task forces

next generation sequencing (and other genomics techniques), proteomics, metabolomics, biobanks, interoperability.

**NEXT GENERATION SEQUENCING**

Solutions emerge when knowledge is shared

The Next Generation Sequencing (NGS) Task Force is one of the five task forces within BioAssist. Each task force tackles bioinformatics problems on a level that surpasses individual groups, but that are (becoming) relevant for the overall research community.

Coordinator Leon Mei starts with an example of their activities. “There are many tools available for aligning sequence fragments. We set up a benchmarking project to compare tools.” NBIC Faculty member Victor Guryev, scientific adviser to the task force adds: “Which tool works best depends on experimental setup, the type of data and analysis required, etc. For an individual researcher, it is very difficult to try out all the possibilities.”

Another result concerns the building of Galaxy-based pipelines. Mei explains: “The Galaxy system has become the de facto platform for sharing and carrying out NGS data analysis. We can now build and deliver tools that everyone can use, and benchmarking is performed automatically.” The latter is done using the researcher’s own data, Guryev adds. “We use a data-centred approach for benchmarking. You select the tools that match the data you want to analyse.”

Both emphasise that the task force’s success goes beyond technical solutions. Mei: “To me, the biggest achievement is that we have built a very active community. In the beginning there were only four; we now have twenty to thirty people participating in the monthly meetings, all willing to share knowledge and problems.” Guryev agrees: “The people are the most important asset. Without them working together, NGS would not change anything.”

**“To me the biggest achievement is that we have built a very active community”**

No one is clever enough to make a difference all on his own

**Nothing New** NBIC’s achievements are gaining international recognition, Mons affirms. “NBIC has created a unique bioinformatics community here and our efforts are being closely watched abroad. For example, Canada, Finland and Switzerland are eager to copy our approach.” He emphasises that it takes strategy and dedication. “What we show with BioAssist is that building such a community is a skill.”

**Numbers based on NBIC Mid Term Review Documentation, March 2011**

**Facts & Figures**

**Support**

- **25** Scientific programmers
- **10** Postdoc positions
- **5** Dedicated task forces

next generation sequencing (and other genomics techniques), proteomics, metabolomics, biobanks, interoperability.
Developing algorithms and analysis methods is a key activity for researchers in bioinformatics. In many cases, however, additional software engineering is needed to turn these 'prototypes' into robust, user-friendly tools. This is where Freek de Bruijn and his colleagues from the BioAssist Engineering Team (BET) step in.

To illustrate what the BET can contribute, de Bruijn refers to Pindel, by now an internationally well-known data analysis program developed by Kai Ye (LUMC), which can detect indels (insertions and deletions) at high resolution in the short reads generated by next generation sequencing. “When Pindel was first developed, basically it boiled down to making data interpretable so that the biologists could get to work,” Marcel Kempenaar of the BioAssist Research Support team (BRS) explains modestly. But thanks to his work, Jos Raaijmakers’ group at Wageningen University can now dig into their treasure trove of PhyloChip-data on the beetroot rhizosphere – the interface between plant root and the soil-based microorganisms. Kempenaar explains: “They processed everything by hand, mostly using Excel, which became very repetitive and time-consuming, and which is only manageable when you know what you’re looking for. The researchers asked us to help them out with developing a more efficient method of working with their dataset because they knew it was a goldmine of new information.”

While talking to the ‘client’, the idea of using Galaxy soon came up. “We created two programs for the group. One to apply a filter to the data and generate subsets and another that delivers the analysis report as a PDF document, including graphics and visuals that are suitable for scientific publications. The client really appreciated our efforts,” Kempenaar affirms, “especially because adjusting the settings is easy, so running a new analysis has become a piece of cake. What at first took them days is now done in just fifteen minutes.”

Ensuring automatic test runs each time an element in the system is adapted is also a BET contribution. “Manual testing gets postponed sometimes or becomes too much work. At that point users lose confidence in the system. With Pindel, testing is always performed, which is very convenient for the user.” Involving the BET brings a fresh perspective, according to de Bruijn, which is needed because developers have often become one with their program. De Bruijn emphasises that they do not interfere with the functionality of a tool. “That’s the one thing we do not touch. What we do is tackle all the other issues that turn a new method into a sustainable and user-friendly tool.”
For centuries mankind has selected plants that completed its needs for food and feed. Varieties were developed that perform as far as quality, yield and agricultural practice are concerned. Such plant breeding practices have been very successful, although complex traits such as yield or drought tolerance have been extremely difficult to improve upon. The revolution in life sciences by molecular biotechnology and genomics has changed the scale and scope of possibilities in plant breeding dramatically. Reconstructing the genome from the billions of short reads, followed by predicting the function of the genes and other regions of DNA in the reconstructed genome lead to identifying key genes and networks and to understanding their functionality. Current plant genomics programmes generate large amounts of data including information on sequences, transcripts, single nucleotide polymorphisms, indels, pathways, gene functions, etc. An important success factor of genomics research is the development of fast and reliable computer tools and software systems allowing management and analysis of the ‘omics’ data. Bio-IT combined with rapidly increasing knowledge in plant genetics will provide a sustainable molecular genetic response to the need for developing varieties with high quality and high nutritional value.

**PLANT BREEDING IN THE GENOMICS/BIOINFORMATICS AGE**

“Whereas in the past, breeders were provided with useful but anonymous molecular markers, the combination of genetic mapping, genome sequencing and bioinformatics will assuredly lead to the unravelling of genes and genetic networks that underlie important agronomical traits. Bioinformatics will support quantitative and molecular genetics to make predictions of combining ability and rational design of new varieties, and will turn plant breeding from tradition into science.”

Rob Dirks (Rijk Zwaan)

**GENOME SEQUENCING**

“Next generation sequencing and “Sequence data in a genome project are increasingly easy to produce in very large volumes with our next generation sequencing technologies. But once produced, the more challenging work of such projects begins. The genome information needs to be structured in such a way that the biology that emerges from the sequence can be seen and analysed. All this is work for and by bioinformaticians. A genome sequencing project such as the one for the potato has therefore primarily been a ‘bioinformatics project’. Now it provides an invaluable basis for biological research and breeding applications.”

Roeland van Ham (Keygene)
“I’m a person who does not like to see a sharp divide between computation and biology. There aren’t two camps in bioinformatics; computer science and statistics are natural partners of biology. Back in the 1950s the basis for statistics was laid by people like Ronald Aylmer Fisher, Henry Mann and Frank Wilcoxon; they all were inspired by biological problems. However, we had some twenty-odd years since the 1970s, when molecular biology came up, in which you could get away without doing maths. As a result, biology had to reinvent its quantitative ridge in the past decade. But in essence biology is a natural, quantitative science, history tells us so.”

“Computers are totally non-creative things. Therefore, computational science is not often seen as a highly creative science, but it is exactly the creativity that attracts me. That creativity lies in the underappreciated ability to ask the right question of a very large dataset. Bioinformaticians often ask very straightforward questions like: What is the genome of this organism? This results in very useful, straightforward answers. But sometimes you need to approach a problem from the side to get into it, because asking the question straight out does not work. You need to figure out how to turn the dataset around to be able to get your answer. That to me is the challenge.”

**SKILLED PEOPLE**

“Computers are not getting faster quickly enough, and disks are not getting smaller quickly enough, but the real bottleneck in bioinformatics is skilled people. There is still this idea in biology that you do not need mathematics, which results in students who have not had maths for five or six years before they get into bioinformatics. Making up these arrears is almost impossible. I often see physicists who do very well in bioinformatics. Since there are not enough jobs for good physicists, I think we should exert the effort to tempt them into bioinformatics. In the meantime, we should of course make sure that every biology student has a firm knowledge of mathematics.”

“Bioinformatics is becoming more the bottleneck of biology, not only because it has become essential to almost any biological experiment, but also because the costs of sequencing are falling at an incredible pace. Five years ago, sequencing a genome or determining SNPs would take the greater part of your budget; the costs of bioinformatics were relatively small. Today, it is at least fifty-fifty, if not going the other way.”

**DUTCH WAY**

“I think there is a very Dutch way of doing science, everything is organised and done in consortia. I see computational biology in the Netherlands as one entity. But if you look closely you also see the little fights and competitions, like everywhere else. But, it is very successful and productive; I see good genomics coming out of it. A consortium provides some critical mass, which is very important in setting up good informatics training programmes in biology, for truly integrating computational biology in biology. We need that more than anything else. I expect that the number of bioinformaticians needed will increase twentyfold over the next twenty years.”

**Ewan Birney, Senior Scientist at the European Bioinformatics Institute (Cambridge, UK), Co-founder of Ensembl**

“History teaches us that bioinformatics is a natural science”

**TOPIC: INFORMATICS AND BIOLOGY AS NATURAL PARTNERS**
Bioinformatics has changed considerably. It was started by a small group of scientists who were actually looked at with some disdain by the experimental community. Soon the experimentalists started to look at bioinformatics as a kind of service for their particular needs. Today we are beginning to see true integration. I believe that there will be no such thing as a bioinformatician in future. A biologist will be trained in doing computing, just as he or she is trained in doing experimental techniques. Of course people will still specialise, but we are promoting this idea of cross-training."

“If I were to receive a large fund with no conditions today, I would definitely put the money into translational medicines, into a project that truly crosses the scales from molecule to patient. That means collecting expression profiles before and during treatment, and looking at genotypes and phenotypes. I want to put the whole story together. Bioinformatics is at the moment in time that we can achieve this, but it requires a lot of effort and money.”

"Today, bioinformatics can put the whole story together"

PHILIP BOURNE, PROFESSOR OF COMPUTATIONAL BIOLOGY
(UNIVERSITY OF CALIFORNIA, SAN DIEGO) | ASSOCIATE DIRECTOR OF THE RCSB PROTEIN DATA BANK | CO-FOUNDER AND EDITOR-IN-CHIEF OF PLoS COMPUTATIONAL BIOLOGY
NBIC Toolbox

The tools shown present a few of the available software packages and computer projects created by NBIC, or in collaboration with NBIC. The growing list is updated on a regular basis and is available at http://tools.nbic.nl

Within NBIC’s bioinformatics research programme BioRange methods are being developed to extract useful information from biological data in order to find the answers to biological questions. Various software packages prove useful for a wider community. These tools are therefore further developed and made publicly available. At the same time, both researchers and scientific programmers are working together in the support programme BioAssist to create workflows for data analysis. Which tool works best depends on experimental set-up, the type of data and analysis required. Various tools are professionalised and turned into sustainable and user-friendly applications. New tools are also developed to complete the data analysis pipeline. NBIC stimulates sharing of tools and experience to avoid duplicating development efforts. All tools are publicly available and everyone is invited to use them. Any feedback is very welcome and will be used for further professional development.

OntoCAT
OntoCAT provides high level abstraction for interacting with ontology resources including local ontology files in standard OWL and OBO formats (via OWL API) and public ontology repositories: EBI Ontology Lookup Service (OLS) and NCBO BioPortal. Each resource is wrapped behind easy to learn Java, Bioconductor/R and REST web service commands enabling reuse and integration of ontology software efforts despite variation in technologies.

LysNDeNovo
LysNDeNovo analyses ETD spectra which utilises the presence of a single fragment ion series to assign the peptide sequence.

Warp2D
Warp2D is an efficient, fundamentally new approach to correct for non linear retention time shifts between complex proteomics and metabolomics LC-MS data sets. Warp2D operates on peak lists and uses the integral of overlapping peak volume of the reference and sample chromatograms in benefit function with Correlation Optimized Warping algorithm.

RoVar
Robust Variant detection in genome sequences using Next Generation Data from various platforms.

bioExpert
The bioExpert project is working towards an environment that allows the creation of high-quality knowledge bases about specific domains of biology.

SequenceHarmony
An entropy-based method, which accurately detects subfamily specific functional sites from a multiple sequence alignment. The algorithm implements a new formula, able to score compositional differences between subfamilies in a simple manner on an intuitive scale.

OntoCAT provides high level abstraction for interacting with ontology resources including local ontology files in standard OWL and OBO formats (via OWL API) and public ontology repositories: EBI Ontology Lookup Service (OLS) and NCBO BioPortal. Each resource is wrapped behind easy to learn Java, Bioconductor/R and REST web service commands enabling reuse and integration of ontology software efforts despite variation in technologies.

LysNDeNovo analyses ETD spectra which utilises the presence of a single fragment ion series to assign the peptide sequence.

Warp2D is an efficient, fundamentally new approach to correct for non linear retention time shifts between complex proteomics and metabolomics LC-MS data sets. Warp2D operates on peak lists and uses the integral of overlapping peak volume of the reference and sample chromatograms in benefit function with Correlation Optimized Warping algorithm.

RoVar Robust Variant detection in genome sequences using Next Generation Data from various platforms.

bioExpert The bioExpert project is working towards an environment that allows the creation of high-quality knowledge bases about specific domains of biology.
The G-Protein Coupled Receptors DataBase (GPCRDB) is a molecular-class information system that collects, combines, validates and stores large amounts of heterogeneous data on GPCRs. The GPCRDB is designed to be a data storage medium, as well as a tool to aid biomedical scientists with answering questions by offering a single point of access to many types of data that are integrated and visualised in a user-friendly way.

www.gpcr.org/7tm

ToPos
The Token Pool Server (ToPoS) is a REST web service that supports distribution of large computational tasks on distributed systems. https://topos.grid.sara.nl/4

WikiPathways
WikiPathways is an open, collaborative platform dedicated to the curation of biological pathways. www.wikipathways.org/index.php/WikiPathways

PathVisio
PathVisio is a tool for displaying and editing biological pathways. www.pathvisio.org

LOFT
LOFT (levels of orthology) can be used to describe the multi-level nature of gene relations. https://trac.nbic.nl/loft

CytoscapeRPC
A plugin for Cytoscape which allows users to create, query and modify Cytoscape networks from any programming language which supports XML-RPC. https://trac.nbic.nl/cytoscapecpc

Molgenis
An collaborative open source project on a mission to generate great software infrastructure for life sciences research. Each app in the MOLGENIS family comes with rich data management interface and plug-in integration of analysis tools in R, Java and web services. www.molgenis.org
Scientists are usually the first to point out the importance of education for research. But when research budgets are allocated, the same scientists are usually not inclined to dedicate part of their newly acquired funds to that education. NBIC decided otherwise and made education an essential element of the overall strategy. Not so much as a matter of principle, but simply fuelled by demand. The field desperately needed bioinformaticians. And life scientists needed to learn about bioinformatics as well, whether they wanted or not. Heringa: “Biologists don’t like mathematics, but now they should no longer ignore it.”

**VARIETY** “There was just not enough bioinformatics education,” says van Gelder. “It was clear that we had to do something.” That is an understatement. BioWise has developed into a large-scale programme that provides highly advanced courses for PhD students, but also introduces bioinformatics to high school pupils and their teachers. According to Heringa, this broad scope comes with the territory. “There are so many different groups of bioinformaticians, ranging from biologists and medical researchers who use bioinformatics tools to hardcore computer scientists whose work takes place in a biological context.” The way bioinformatics education is currently organised in the Netherlands represents and further contributes to that heterogeneity. There are dedicated bioinformatics programmes at universities (MSc) and universities of applied sciences (‘HBO’, BSc); but bioinformatics is also offered as a specialisation within informatics, biology or life sciences programmes. Heringa: “That results in different types of bioinformaticians; some are more informatics-minded, others lean more towards the biology side. And that is fine, because we need that variety.”

“Biologists should no longer ignore mathematics”

**SLOW PROCESS** By targeting all levels, from high school students to professionals, BioWise also aims to stimulate an adequate flow, van Gelder adds. “Having the right bioinformatics PhD students implies that there needs to be adequate education on the master and bachelor levels, which in turn means that we need the right curriculum in high schools. This is a continuous process because the field is developing so fast.” She herself noticed that the BioWise activities are showing results. Van Gelder teaches bioinformatics at the Radboud University Nijmegen. “The other day I had some students who could easily perform the assignments. When asked, they told me they had been introduced to bioinformatics in high school and had also participated in our bioinformatics@school workshop. It is really great to hear that our initiatives increase the level of knowledge in this group.”

Changing educational practices can only be done by involving teachers, not only in high schools, but also at the universities of applied sciences. Van Gelder explains:
“At high schools, our primary goal is to convince teachers, mostly biology teachers, that bioinformatics is here to stay and that it is more than just a side topic. To a large extent already, bioinformatics is biology.” Getting teachers on board ensures a lasting effect: students go, teachers stay. Heringa: “Changing educational programmes is an awfully slow process. The processing speed of the educational system can never keep up with the pace of scientific and technological developments. You really need this community to get things going.”

“We need to convince teachers that bioinformatics is here to stay”

**BOTTOM-UP** Their efforts in mobilising the community are paying off, as BioWise is increasingly being approached by those in the field for help. This is exactly the mechanism that Heringa and van Gelder strive for. “We do not develop courses ourselves; the field should take the lead and we can offer support in all kinds of ways,” says van Gelder. BioWise is not only gaining ground in the Netherlands. Internationally, the NBIC approach is increasingly being referred to as a shining example. Heringa: “The Mid-Term Review committee stated that ‘NBIC’s educational programme is phenomenal’. We clearly have succeeded in setting up a coherent programme that is more than a collection of individual courses.” Naturally, the question of why BioWise has become such a success pops up. Heringa thinks it is their bottom-up approach. “We do not dictate the activities of others. We can push and help and stimulate, but in the end, the field has to do the work.” He also feels that the common need has proved crucial. “Everyone acknowledges the need for bioinformaticians and the fact that no one can tackle this need alone.”

Looking for advanced courses in bioinformatics? The NBIC PhD School’s programme covers a variety of topics and techniques taught by experts with hands-on experience. Dick de Ridder, assistant professor at Delft University of Technology, is one such expert who teaches two courses: Pattern Recognition (with Perry Moerland and Lodewyk Wessels) and Algorithms for Biological Networks (with Marcel Reinders and Anton Feenstra).

“In the Pattern Recognition course, we want to make the students aware that the algorithms they use are not just ‘black boxes’, but that they are based on certain assumptions and that parameters can be adjusted to generate a different result.” The course discusses basic techniques, some of which have been in use for a long time. De Ridder explains: “This is not so much about presenting the state of the art, but about stimulating a critical mindset with the students.”

The ‘Algorithms for Biological Networks’-course is more focused on the state of the art. “In this course we present the latest techniques and their specific applications for deriving biological information from the data. With this course, we hope to inspire the students to apply new techniques in their research and to be creative.”

The participants range from biologists with a superficial knowledge of statistics to computer experts who are less familiar with the biological context. “Dealing with this variety is a challenge, but I think we have found the middle course. And the mutual interaction between the participants is a great way for them to learn from each other.”

“We hope to inspire the students to apply new techniques in their research and to be creative.”

---

**Facts & Figures**

**Education**

- **10** Advanced PhD courses (co)organised by NBIC
- **15** Events organised by RSG Netherlands
- **22** Teacher trainings
- **11,032** High school students participated in bioinformatics practical
The bioinformatics@school programme targets high school students and their teachers. “Bioinformatics is still unknown to most people,” says Hienke Sminia, education officer and coordinator of the programme. “Younger generations in particular will encounter bioinformatics applications, for example in health care. It is important that they learn about the role and contribution of bioinformatics.”

One of the ways to introduce high school students to bioinformatics is the hands-on workshop ‘Bioinformatics: a bit of life’, one of the so-called DNA Labs on the Road. Since its introduction in 2006, more than 400 classes involving almost 12,000 students have participated in this workshop. Sminia explains: “Bioinformatics is really well-suited for educational applications because it is very accessible. And it is a great way to introduce students to science.”

Not only are the students brought into contact with bioinformatics, but their teachers are as well, which is probably even more important. “The bioinformatics @school programme also covers specific activities for teachers, such as training sessions for biology teachers in which they learn how to prepare lessons using bioinformatics tools.”

At the moment, Sminia and colleagues are developing an augmented reality module to create a ‘3D protein scanner’ to be used in schools. “Using a webcam, a student can for example scan a carton of milk to find out which proteins are present.” Sounds like fun. “That is also an essential element in education,” says Sminia. “When students can relate to a topic and enjoy thinking about it, they process new information much better.”

“Bioinformatics is really well-suited for educational applications because it is very accessible.”

Herself teaches bioinformatics at the UAS Leiden. “But it is also a good instrument for getting to know each other. Bioinformatics is still a small field. Being able to talk with others concerning your work and to check whether you are on the right track is very helpful. We also see the emergence of the first real collaborations on developing new modules and teaching materials together. Without LOBIN this would not have happened. Collaborating is important because bioinformatics is such a fast-moving field; it is impossible to keep up all on your own.”

Being connected to NBIC ensures access to the broader bioinformatics field and to the latest scientific developments, which are employed in LOBIN’s ‘teach-the-teacher’ sessions to bring bioinformatics teachers up to speed. A new activity is a career event for bachelor students. Krosenbrink: “This fits the development of bioinformatics as a mature field. Such events have been around in other areas for a long time.”

“We see the emergence of the first real collaborations on developing new modules and teaching materials together.”

Three Universities of Applied Sciences (UAS) in the Netherlands offer bachelor programmes in bioinformatics. Together they form the so-called landelijk HBO-overlag network, LOBIN, which is chaired by NBIC. LOBIN activities relate to exchanging curriculum information and teaching materials, organising teacher training and promoting the bachelor programmes to prospective students. UAS that offer bioinformatics as part of a life sciences or informatics curriculum can also participate in LOBIN activities.

“Currently, LOBIN is primary focused on the educational content,” says Marja Krosenbrink, secretary of LOBIN. She laughs: “After all, we are teachers.” Krosenbrink
Nowadays eating is not so much a matter of survival till tomorrow as it is a matter of reaching old age while remaining healthy. Therefore food and nutrition science increasingly focuses on research into harnessing food for health and preventing food-related diseases. By studying how food components are digested, absorbed, metabolised and utilised, their effects on genes, cells, organs and the whole person can be understood. Nutrigenomics, metagenomics, metaproteomics, metatranscriptomics and metabolomics approaches are essential to investigating the complex interactions between food components, microbiota and the host. Such studies require strong multidisciplinary cooperation, and bioinformatics is becoming increasingly important. The opportunities are numerous and may range from facilitating access to tailored IT-tools to full standardisation of nutrition experiments and migration of experimental data into databases. Ultimately, bioinformatics will pave the way to translating experimental results into knowledge about the relationship between food, diet and health. This will significantly accelerate the ability of investigators to discover the essential implications of nutrients and food components and to study the complex metabolic interactions underlying human health and disease.

**SYSTEMS BIOLOGY APPROACH**

“Nutrition science is extremely complicated because the effects of diet on health are very subtle and involve multiple processes on all levels of (molecular) physiology. In other words: a real systems biology approach is essential. Nutrigenomics is causing a real conceptual shake-up of nutrition research, with bioinformatics at the basis of this.”

Ben van Ommen (TNO, NuGO)

**MULTI-FACETED AND COMPLEX**

“Both from industry as well as from the scientific community there is growing interest in the contribution of our diet to health. However, the impact of nutrients on our body is multi-faceted and interactions between nutrients, microbiota and human cells are complex. Thanks to bioinformatics we have started to unravel this complexity and are at the beginning of a more systemic understanding of the role of food and nutrition in health.”

Jan Sikkema (formerly TI Food & Nutrition, now UMC Groningen)
PhD theses

16 Automatic sign language recognition inspired by human sign perception
Gineke ten Holt
Delft University of Technology, 23 June 2010
Promotores: Prof. dr. M.J.T. Reinders, Prof. dr. H. de Ridder

17 Comparing building blocks of life: sequence alignment and evaluation of predicted structural and functional features
Walter Pirovano
VU University Amsterdam, 15 January 2010
Promotor: Prof. dr. J. Heringa | Co-promotor: Dr. K.A. Feenstra

18 Phenotype-guided disease investigation using bioinformatics
Martin Oti
Radboud University Nijmegen, 23 April 2010
Promotores: Prof. dr. H.G. Brunner, Prof. dr. M.A. Huynen

19 Exploiting noisy and incomplete biological data for prediction and knowledge discovery
Yunke Li
Delft University of Technology, 7 October 2010
Promotor: Prof. dr. M.J.T. Reinders | Co-promotor: Dr. D. de Ridder

20 Do you know what i know? Situational awareness of co-located teams in multidisplay environments
Olga Kulyk
Twente University of Technology, 14 January 2010
Promotores: Prof. dr. A. Nijholt, Prof. dr. G.C. van der Veer
Co-promotor: Dr. E.M.A.G. van Dijk

21 X-ray structure re-refinement. Combining old data with new methods for better structural bioinformatics
Robbie Joosten
Radboud University Nijmegen, 12 May 2010
Promotor: Prof. dr. G. Vriend

22 Gesture interaction at a distance
Wim Fikkert
Twente University of Technology, 11 March 2010
Promotores: Prof. dr. A. Nijholt, Prof. dr. G.C. van der Veer
Co-promotor: Dr. P.E. van der Vet

23 Service-oriented discovery of knowledge: foundations, implementations and applications
Jeroen de Bruin
Leiden University, 18 November 2010
Promotor: Prof. dr. J.N. Kok

24 Personalised access to social media
Maarten Clements
Delft University of Technology, 6 December 2010
Promotores: Prof. dr. M.J.T. Reinders, Prof. dr. A.P. de Vries

25 A picture is worth a thousand words. Content-based image retrieval techniques
Bart Thomée
Leiden University, 3 November 2010
Promotor: Prof. dr. J.N. Kok | Co-promotor: Dr. M.S. Law

26 Proteomics screening of cerebrospinal fluid. Candidate proteomics biomarkers for sample stability and experimental autoimmune encephalomyelitis
Theresa Rosenling
University of Groningen, 20 December 2010
Promotor: Prof. dr. R.P.H. Blischke

27 Transcriptome profiling of infectious diseases and cancer in zebrafish
Anita Ordas
Leiden University, 29 June 2010
Promotor: Prof. dr. H.P. Spank | Co-promotor: Dr. A.H. Meijer

28 Small RNA evolution and distribution patterns based on digital gene expression profiling
Samuel Linsen
Utrecht University, 9 February 2010
Promotor: Prof. dr. E.P.J.G. Cuppen

29 Interaction and evolutionary algorithms
Ron Breukelaar
Leiden University, 21 December 2010
Promotores: Prof. dr. T.H.W. Bäck, Prof. dr. J.N. Kok

30 Integrative bioinformatics of metabolic networks
Richard Notebaart
Radboud University Nijmegen, 6 May 2009
Promotores: Prof. dr. R.J. Siezen, Prof. dr. B. Teusink
New drugs are routinely tested on rodents before they can be tested on humans. But in the last ten years many drugs that performed very well in rodents fell short during human clinical trials. Are these animal models really as predictive as we had like to believe? It is Umesh’s job to find out while pursuing his PhD study at the Bioinformatics Laboratory, led by Perry Moerland. He is trying to determine the similarities and differences between human and rodent at the molecular level. “I am well on my way to comparing rodent microarray data to human expression sets,” he says, “we will submit my first article soon.”

“At the beginning it was slow going; you would not believe how many different abbreviations people use for a single disease,” Umesh says. “Most of the work I’ve done so far has centred on building a homogeneous framework that can systematically store data from gene expression studies from different resources such as different microarray chips, experimental protocols and data formats. Because everything else will be built and analysed on the data derived from this database, it had to be absolutely perfect.”

The next step was the development of an R package to enable researchers to use the database for their own specific domain of study. Umesh: “Many studies easily have more than a hundred patients. Add to that the 100-plus microarrays from the rodents, each containing thousands of genes: it can take a while to structure and compare all those experiments one by one. But using my R package, you can download a large number of microarray experiments in a few hours.” The necessary algorithms are almost done; now the program has to be user friendly, which means lots of documentation and examples.

But Umesh is far from discouraged. “I think everything is going according to schedule. And I get a lot of motivation from bouncing ideas with my fellow researchers. The NBIC PhD course I participated in earlier this year was especially inspiring. We had to defend or attack a scientific paper on evolution and comparative genomics every day. Even though it was hard as it was a little outside my area of expertise, it resulted in extremely valuable discussions. It was a real biological addition to my computer science background.”
Bioinformatics is at the base and at the centre of any major advance in medical and molecular biology today, and there are centres for computational biology all over the world. Bioinformatics has succeeded in spanning the distance between method developer and application, between mathematics and biology. That’s a major achievement. Unfortunately, the perceived importance does not yet match its actual impact. For example, bioinformatics is still not institutionalised on the level of university departments.”

“Within five years, I expect physicians to start checking their patients’ genome before prescribing any medicines. Instead of guessing which of the available drugs will treat the disease best, the physician will have this information at his or her disposal through genomic tests. The next step will be truly personalised medicine: therapies especially developed for particular subgroups of patients. Therapies for cancer will be based on individual tumour characteristics and patient features such as metabolism, age and gender. But this will take at least another decade. We don’t have computers that are fast enough for our needs, and we currently don’t have the necessary funds, algorithms, or people to make faster progress.”

**UNIQUE DUTCH APPROACH** “The newer branches of bioinformatics evolved almost separately from existing groups in computational or theoretical biology. In the Netherlands, however, the unique event occurred that funds were showered upon some very unselfish scientists, who, instead of building fiefdoms and fighting each other, bundled all forces in bioinformatics into NBIC, the Netherlands Bioinformatics Centre. A rare occasion, because in bioinformatics competition is quite aggressive, perhaps because one can easily operate alone. But cooperation proves to be advantageous; the Dutch initiative has a truly impressive output, not only in terms of scientific achievements but also in terms of its educational programme. NBIC’s mobile lab has already reached tens of thousands in schools. If all researchers in the world would behave like those who joined NBIC, computational biology would be far advanced into the future.”

**CATCHING UP BY CHINA** “In the 1990s, Europe was in many ways ahead of the USA in bioinformatics and computational biology, and Europe still stands strong, particularly in predicting protein structure. However, the USA started off fast in next generation sequencing, the field in which I expect the largest progress in the coming years. But now China seems to have taken over. China is investing in a mind-boggling manner and pace. The Beijing Genome Institute built an infrastructure in Shenzhen which puts the rest of the world at once in serious, perhaps even impassable arrears. Will this mean unemployment for European bioinformaticians? I don’t know, but the fact that China will dominate data important for the health of all of us is a new prospect.”
“I witnessed the beginnings of bioinformatics and have worked in the field ever since. But today I would say that my specific field of research is biocuration: the organisation of data and knowledge for the life sciences, first with Swiss-Prot and now with nxProt, our new human protein-centric resource. Biocuration is a subfield of bioinformatics which over the past several years has developed into its own discipline. I call it the ‘bread-and-butter’ of bioinformatics as data are essential to any type of bioinformatics research.”

“Our main challenge is to speed up the process of building high-quality knowledgebases. Currently, these are still chiefly built by people - biocurators - reading papers and capturing and summarising the results by typing them into databases, an archaic method when you think about it. If authors could add semantic tags to their papers, the process could be automated. That would speed up the process of biocuration tremendously. Knowledgebases would also become more accurate as the experimentalists themselves would be part of the process. The main barriers to attaining our goal are not technological, but sociological. Semantic tagging requires extra efforts. But there is slow progress. PLOs journals, for example, are encouraging such progress. And in the end, it is just inevitable.”

“Bioinformatics is often called the bottleneck of life sciences. In a way, that is true. Currently, analysing the data obtained after a couple of days of experiments takes considerably longer than those two days. However, it is also the result of a misconception. People expect bioinformatics to be as fast as high-throughput data acquisition, but that would be magic. The technology, the computers, may be fast, but human expertise is always required to produce the actual knowledge. Getting data is fast; getting knowledge out of it is time-consuming.”

“Bioinformatics does not translate directly into benefits for society, but it has become indispensable to the life sciences, as for example in drug development. Many HIV inhibitors could not have been developed without 3D models or sequence alignments. A tool like BLAST has advanced research tremendously. Modern life sciences research has become high-throughput based and requires the help of bioinformatics to make sense of the accumulated results. The life sciences and bioinformatics cannot be separated anymore.”

THE NETHERLANDS ON TRACK
“Europe wide, the Netherlands has always been a well-known player in bioinformatics, although perhaps not specifically in the field of biocuration.

It is, however, a key player in the field of the semantic web. The Netherlands was also one of the first European countries, after Switzerland and Spain, to create a world-class national centre in bioinformatics. Although large centres such as the EBI and NCBI have had a huge impact on bioinformatics in the past 20 years, countries interested in maintaining a competitive edge in science need to support bioinformatics and biocuration. National institutions that provide services and foster education will have the qualified scientists to support modern research.”
31 Graph-based methods for large-scale protein classification and orthology inference  
Arnold Kuzniar  
Wageningen University, 8 November 2009  
Promotor: Prof. dr. J.A.M. Leunissen | Co-promotor: Dr. R.C.H. J. van Ham

32 Inferring the influence of cultivation parameters on transcriptional regulation  
Thao Knijnenburg  
Delft University of Technology, 21 March 2009  
Promotor: Prof. dr. M.J.T. Reinders | Co-promotor: Dr. L.F.A. Wessels

33 Bayesian networks for omics data  
Arond Gavai  
Wageningen University, 8 June 2009  
Promotors: Prof. J.A.M. Leunissen, Prof. dr. M.R. Muller

34 Data mining scenarios for the discovery of subtypes and the comparison of algorithms  
Fabrice Colas  
Leiden University, 4 March 2009  
Promotor: Prof. dr. J.N. Kok

35 Multinomial language learning, investigations into the geometry of language  
Stephan Raaijmakers  
University of Tilburg, 1 December 2009  
Promotors: Prof. dr. A.P.J. van den Bosch, Prof. dr. W.M.P. Daelemans

36 Rational systems in control and system theory  
Jano Nemcová  
VU University Amsterdam, 2 December 2009  
Promotor: Prof. dr. J.H. van Schuppen

37 Webservices for transcriptions  
Pieter Neerincx  
Wageningen University, 14 September 2009  
Promotor: Prof. dr. J.A.M. Leunissen

38 Gesture recognition by computer vision: an integral approach  
Jeroen Lichtenaar  
Delft University of Technology, 13 October 2009  
Promotor: Prof. dr. M.J.T. Reinders | Co-promotor: Dr. E.A. Hendrika

39 Signaling pathways in cancer: a matter of dosage  
Cláudia Gaspar  
Erasmus University Rotterdam, 26 February 2009  
Promotor: Prof. dr. R. Fedde

40 Computational genomics for prokaryotes  
Evert-Jan Blom  
University of Groningen, 11 December 2009  
Promotors: Prof. dr. O.P. Kuipers, Prof. dr. J.B.T.M. Roerdink

41 Spatio-Temporal Framework for Integrative Analysis of Zebrafish Developmental Studies  
Mounia Belmamoune  
Leiden University, 17 November 2009  
Promotor: Prof. dr. J.N. Kok

42 Dynamic software infrastructures for the life sciences  
Morris Swertz  
University of Groningen, 15 February 2008  
Promotors: Prof. dr. R.C. Jansen, Prof. dr. E.O. de Brock

43 Models of natural computation: gene assembly and membranes system  
Robert Brijder  
Leiden University, 3 December 2008  
Promotor: Prof. dr. B. Rozenberg | Co-promotor: Dr. H.J. Hoogeboom

44 Spatio-temporal gene expression analysis from 3D in situ hybridisation/images  
Monique Welten  
Leiden University, 27 November 2007  
Promotors: Prof. dr. S.M. Verduyn Lunel, Prof. dr. H.P. Spaink | Co-promotor: Dr. F.J. Verbeek

45 Comparative Genomics of Eukaryotes  
Vera van Noort  
Radboud University Nijmegen, 8 January 2007  
Promotor: Prof. dr. M.A. Hueyman
Anand Gavai:

“The success of my PhD project was definitely a team effort”

While doing his PhD he discovered his two passions: bioinformatics research and commercial application. He analysed microarrays, won the NBIC Venture award and kick started his commercial career. Why? What does his research matter? “Bioinformaticians have the advantage here,” says Anand, “since they usually produce a tool or program that others can use.” And sometimes the effort grows into a beloved application used by thousands of people, like his PhD project. “However, it definitely was a team effort,” Anand explains. “Without my supervisor at Wageningen University, Professor Jack Leunissen, that would never have happened. At first I didn’t even know he had entered me in the NBIC Venture challenge. He just asked for an abstract; I thought it was for a conference or something. But suddenly I was standing in front of a jury, being grilled about markets and customers and how to transfer academic thoughts into a product.”

During the first two years of his project, Anand built a database that could compare Affymetrix microarrays of all shapes and sizes and analyse the data with Bayesian algorithms. “I was lucky and ended up in an amazing group that actually had other bioinformaticians, a rarity in 2004,” Anand explains. “It was such a new field, half of the time we had no idea whether what we were doing would accomplish anything. This meant that I had a lot of freedom and my opinion was valued.” His unexpected entrance in the Venture Challenge turned out well: he won the 30,000 euro that came with the first prize.

The prize has led to a lot of travelling and presenting the program at conferences. “That was a great time to network,” Anand says enthusiastically. “I can honestly say that I know eighty-five percent of the bioinformaticians in the Netherlands and also quite a large number abroad.” When he finished his PhD, he knew that this commercial vibe was just what he needed. He recently took a job at Agendia where he is working on improving dNa chips that analyse breast cancer tissue samples to predict possible metastization of the cancer. “I feel like I am just having fun. Ninety-nine percent of what I am doing is a lot like my PhD work, with the additional bonus that I have the freedom to explore the most current techniques.”
Gathering mass to make noise

Research findings, computational methods, prototypes of software and databases, educational material and, perhaps most important, human capital – the output of the NBIC network takes many forms. NBICommons, the community valorisation programme of NBIC, helps to create added value from this output. Ruben Kok, managing director of NBIC and Karin van Haren, manager communication at NBIC, explain how NBICommons operates to ensure optimal use of results.

The strong body of expertise that is the NBIC network is the basis of our growing international scientific status,” according to Kok. “But extra effort is needed to create tools with high usability and visibility, and thus added value. NBICommons is about making technology usable and used.” To do justice to all the different activities, NBICommons supports four lines of approach.

The first line concerns dissemination and communication. Kok explains: “We help to bring parties and activities together, to set the development track in motion and to employ the right communication channels.” Van Haren adds: “In the first years, we focused on ‘internal’ communication to make everyone involved in a NBIC project feel part of a broader community.” She further explains: “That is where it starts. Next they share their enthusiasm with others and thus contribute to the visibility and vitality of the NBIC network.”

“We first focused on building the NBIC community”

Concept Web Alliance, which spotlights the Netherlands. Kok says: “And this in turn resulted in our involvement in several European projects, including the Innovative Medicine Initiative.”

GAINING INFLUENCE According to Kok, the added value of all these activities is that the bioinformatics field is gaining influence on the right stages. “By joining forces with other national bioinformatics centres in Europe you can start to make some noise, so to speak, and influence processes that are important to Dutch bioinformatics.” Profiling NBIC on the international stage is also a key objective of NBIC’s communication strategy. Van Haren explains: “We noticed that NBIC is becoming internationally known. This recognition is strongly fuelled by the fact that our community really has something to offer. Promoting yourself only makes sense when you can substantiate your message with actual results and achievements.”
The NBIC calendar is filled with meetings and events, but there is one occasion where the whole community comes together: the annual NBIC Conference. And its development nicely mirrors the development of NBIC itself. “At first the conference focused on communicating scientific results from NBIC research projects,” says Femke Francissen, communications officer at NBIC. “Now the conference presents a broader range of topics and activities and is increasingly attracting satellite meetings.” The Regional Student Group (RSG), for example, takes the opportunity to organise a retreat prior to the conference, and during the 2011 edition the Bioinformatics Industrial User Platform (BIUP) hosted a meeting with industrial representatives.

With approximately 250 participants each year, ranging from PhD students to established scientists and from biomedical researchers to software engineers, the conference proves interesting to a broad group. “There is a real community atmosphere where people know each other; yet our evaluations show that each year participants also establish new contacts. There is ample social interaction and people actively seek out each other’s company.”

The conference programme is as diverse as the community itself. Francissen: “We make sure there is something interesting for everyone. So we offer the ‘classic’ scientific lectures and poster sessions, but also an application showcase, where software developers can demonstrate their work. Participants can get up to speed with specific tools during our software tutorials. And there is always a ‘fun’ workshop to attend, like this year’s ‘Informaticians are from Mars, biologists from Venus’.”
B ringing different parties with a shared interest together in an informal setting to enable making new contacts and exchanging information is a well-known approach to networking and community building. The Bioinformatics User Platform, better known as BIUP, is no exception, according to Marco de Groot of the BIUP organising team and scientist bioinformatics at the DSM Biotechnology Centre. “Quite a number of companies in the Netherlands use bioinformatics, but there was no exchange of common efforts,” de Groot explains. About two years ago, an initiative was started to organise informal meetings.

NBIC offers logistic and administrative support, which is much appreciated by the BIUP team. De Groot says: “NBIC appreciates the interest from industry, while companies cherish the opportunity to get involved and share their views. Although BIUP is an independent initiative, there is a fruitful overlap with NBIC’s activities.”

For instance, the program extracts mutation data from literature and sometimes needs to scan more than 100,000 relevant papers. The system understands the content and ranks all data according to the customer’s interest.

With Bio-Prodict, Joosten is building on his Master thesis, which he completed in Gert Vriend’s group (CMBI, Nijmegen). He developed his idea during his PhD research at Wageningen University. The links with NBIC are still strong. “We closely collaborate with Gert Vriend and new developments, which are sometimes sponsored by NBIC, are incorporated into our operations. NBIC sponsored our website and gave us the opportunity to present ourselves at important events.” The key to Bio-Prodict is that they accelerate research, says Joosten. “Our customers have their database within a couple of weeks and can extract relevant information within minutes, compared to years of work when generating such systems manually.”
Industrial biotechnology

Industrial biotechnology, also known as white biotechnology, implies bio-based industrial processes. It uses microorganisms, whole cells or cell components such as enzymes, to generate a broad range of industrially useful products. Some examples are renewable chemicals, biofuels, pharmaceutical intermediates and food processing enzymes. Modern techniques like DNA sequencing, gene expression and protein engineering are often applied to optimise cell cultures and enzymes. Bioinformatics plays a role in the upfront design of such bioconversion systems and also afterwards in the characterisation of the resulting strains and cells. For instance, it provides the ability to read genomes using (re)sequencing and to write DNA using gene synthesis. Genomics (transcriptomics, proteomics and metabolomics) are important emerging technologies for characterising biosystems on a detailed level. Computational techniques lead to comprehensive knowledge on how cellular components like genes, proteins and metabolites are regulated or how microorganisms interact in a complex starter culture. Unravelling such mechanisms at the molecular level may lead to more sustainable production processes and optimisation of process conditions. Bioinformatics provides a promising approach to generate entirely new insights for creating innovative industrial biotechnology.

PERFORMANCE OF MICROORGANISMS

"Many Kluyver Centre research projects target linking the performance of microorganisms in industrial processes to their genetic make-up. The 1000-dollar genome, an iconic target in human genetics, is already a reality for microorganisms – but availability of the primary sequence data is only a start. Intensive collaboration with NBIC scientists – some of whom spent part of their time at the Kluyver Centre – has been vital in setting up efficient in-house pipelines for the interpretation of genome sequence data."

Jack Pronk (Kluyver Centre)

GETTING THE MOST OUT OF STRAINS

"Bioinformatics and modelling have become essential disciplines at DSM’s Life Sciences cluster to develop the next generation of microbial production strains and bioproducts. Design practices are successfully applied at multiple cellular levels. For example, DNA redesign, protein engineering and metabolic pathway engineering are successfully applied to get the most out of strains. Moreover, genomics data integration, modelling and visualisation are key for iterative target selection in these programmes."

Hans Roubos (DSM)
Methods for analysing genetic association studies. Application to cardiovascular disease
Olga Souverein
University of Amsterdam, 5 April 2007
Promotor: Prof. dr. A.H. Zwinderman | Co-promotor: Dr. ir. M.W.T. Tanck

Operating characteristics for the design and optimisation of classification systems
Thomas Landgrebe
Delft University of Technology, 19 December 2007
Promotor: Prof. dr. M.J.T. Reinders | Co-promotor: Dr. R.P.W. Duin

Familial colorectal cancer, omics and all that jazz
Joanna Cordoso
Leiden University, 14 February 2007
Promotores: Prof. dr. R. Fodde, Prof. dr. J. Morreau
Co-promotor: Dr. J. Boer

Pharmacophylogenomics - Explaining interspecies difference in drug discovery
Tim Nulsen
Radboud University Nijmegen, 14 September 2007
Promotor: Prof. dr. J. de Vlieg | Co-promotor: Dr. P. M. A. Groenen

Proteomics of Body fluids
Lennard Dekker
Erasmus University Rotterdam, 10 October 2007
Promotores: Prof. dr. P. A. E. Sillevis Smit, Prof. dr. C. H. Bangma
Co-promotores: Dr. T. M. Luidens, Dr. ir. G. W. Jenster

Affect and Learning: a computational analysis PhD thesis
Joost Broekens
Leiden University, 18 December 2007
Promotor: Prof. dr. J. N. Kok
Co-promotores: Dr. F. J. Verbeek, Dr. W. A. Kosters

Mathematical aspects of infectious disease dynamics
Barbara Boldin
Utrecht University, 5 September 2007
Promotores: Prof. dr. G. Diekmann, Prof. dr. M. J. M. Bonten

Computational genomics of gram-positive bacteria
Jos Boekhorst
Radboud University Nijmegen, 23 May 2007
Promotores: Prof. dr. R. J. Siezen | Co-promotor: Prof. dr. M. Kleerebezem

On the quality of NMR structures. Methodology and tools for NMR data and structure validation
Sander Nabuurs
Radboud University Nijmegen, 9 February 2006
Promotor: Prof. dr. G. Vriend | Co-promotor: Dr. G. W. Vuister

Statistical methods for microarray data
Jelle Goeman
Leiden University, 8 March 2006
Promotores: Prof. dr. J. C. van Houwelingen, Prof. dr. S. A. van de Geer

The nuclear receptor ligand-binding domain: from biological function to drug design. A protein family-based approach
Simon Folkertsma
Radboud University Nijmegen, 3 November 2006
Promotor: Prof. dr. J. de Vlieg | Co-promotor: Dr. P. I. van Noort

From sequence to structure and back again
Victor Simosia
VU University Amsterdam, 7 July 2005
Promotor: Prof. dr. J. Heringa

Experimental DNA Computing
Christiaan Henkel
Leiden University, 23 February 2005
Promotores: Prof. dr. H. Spank, Prof. dr. G. Rozenberg, Prof. dr. T. Bäck

The last mile of the protein folding problem: A pilgrim’s staff and skid-proof boots
Elmar Krieger
Radboud University Nijmegen, 27 September 2004
Promotor: Prof. dr. G. Vriend
Jules Kerssemakers: “It feels good to be part of a large community”

He has stuck to his home town of Nijmegen. Jules stayed to study molecular life sciences and again to do his PhD. “In high school I liked chemistry, biology and physics the most. Studying in Nijmegen fitted that the best since they approach life sciences from a more (bio)chemical point of view,” he explains. From childhood on Jules has been ‘fiddling’ with computers, a hobby which combines well with biology in the discipline of bioinformatics. His talent was noticed by Professor Gert Vriend of the Modelling and Data Mining group at the CMBI during some courses Kerssemakers followed for his bachelor’s degree. “And after an internship with him, he found me interesting enough to offer me a job. I didn’t refuse. So I stayed in Nijmegen.”

For his PhD project Jules is working on new methods and computer programs that should make information now known to just a few scientists easily accessible to other researchers. “This concerns all kind of information about proteins. I use my biological knowledge to teach the computer how to solve biological questions. I do the thinking and the computer does the handiwork.” In the meanwhile he has been following various courses at the NBIC PhD School, such as information management, statistics and profile recognition. “It is possible to follow all the available courses while doing your PhD. You are not obliged to, but doing so is extremely worthwhile. In one week you are taught the state of the art of that specific topic by top scientists, and you meet other PhD students and researchers from around the country. This is the beginning of your future network,” says Jules, who is also an active visitor to the BioCafe’s organised by the Regional Student Group.

He speaks even more enthusiastically about the annual bioinformatics meeting: “When you are at the conference you realise that you’re part of a large community. There all bioinformaticians from the Netherlands and even several from abroad attend. That is very important for us since we have in our country as many bioinformaticians as there are just on the University of California San Francisco campus. Therefore it is necessary to organise ourselves and to participate in the world scene. NBIC plays the role of super campus of the Netherlands. That is the real value of our bioinformatics network organisation.”

Jules Kerssemakers started his PhD study in 2009 at the Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Nijmegen.
Similar souls

“W e have become the organisation for bioinformatics PhD students to meet, discuss research ideas and form an informal network in the field, independent from their PI’s, of course.”

Enthusiasm for the Regional Student Group (RSG) Netherlands is contagious when talking to its secretary Miranda Stobbe. “We are currently moving towards the third generation of board members and our events have a decent turnout. I can in all modesty say that I am proud of what we have accomplished so far.” Miranda, along with founding mother and former President Jayne Hehir-Kwa, has been with the Dutch RSG since its beginning. The founders saw the dire need for a student network. “Bioinformatics is such a new field that a lot of us aren’t in bioinformatics research groups,” Jayne explains. “For example, I worked in the Human Genetics department at Nijmegen. All my colleagues were molecular biologists and clinical geneticists, so it was great finding an external group to use as a sounding board for bioinformatics problems. With the RSG we try to bring people together, because it is so much easier to ask for an opinion on a problem when you know a similar soul.”

INTERNATIONAL NETWORK The RSG Netherlands is embedded in an international network of 21 RSGs, each of which tries to fulfil the needs of the local PhD students. Most create networking and study opportunities, either in person or online. Miranda explains: “We have the advantage that our country is small enough to convince people to take a train and come to a workshop. That’s a lot harder in India or South-West Africa. And of course we are supported by NBIC, not just financially, but they also help us with promoting and organising events.”

In 2008, NBIC initiated the Regional Student Group (RSG) – a group of bioinformatics PhD students in the Netherlands – which is part of the worldwide network of RSGs coordinated by the Student Council of the International Society for Computational Biology, ISCB. The RSG aims to initiate and stimulate scientific discussion as well as collaboration between PhD students and young scientists in the field of bioinformatics and computational biology.

Since the foundation of RSG NL in 2008 there have been workshops on soft skills, bioinformatics pub quizzes, informal lectures and company visits. Their speed dating concept was so successful that organising committees of international conferences like ISMB and ECCB have copied the idea. “But we still do it too,” elaborates Miranda. “It’s usually the evening before a conference, and everyone has just one minute to explain his or her research to the other person. That way you both get to practice your elevator pitch and you already know people to talk with during the conference.”

HIGHLIGHT The keynote dinner the evening before the NBIC conference 2011 was a highlight; a group of PhD students had pancakes with the editor-in-chief of PLoS Computational Biology, courtesy of the RSG. “It is so much easier to talk to someone like that with just us PhD students present. No PI’s who engage in discussions that have been going on for five years, or who we need to impress. Incidentally, that’s our only hard rule: everyone is welcome, except PIs. Let’s be honest, when your boss is at a party it’s hard to just relax and enjoy.”

In the future RSG NL wants to involve more PhD students, and even master students, in their activities. And they are always looking for new board members!
Bioinformatics Network

Like a spider in its web, the NBIC consortium fulfils its enabling role as a cross-institute centre which leads to a true network.

The open organisation facilitates setting up collaborations with Dutch Life Sciences research consortiums as well as informatics and e-science research initiatives. Moreover, the network paves the way for international collaborations with sister organisations and global initiatives.

Unravelling the complexity of life by the integration of bioinformatics-based technologies.

The growing (inter)national network is excellently geared to be a home to and a marketplace for a wide variety of bioinformatics and engineering projects. The interweaving gives a good overview of the many bioinformatics projects.

Number of bioinformatics projects taking place at the intersection of life sciences research and enabling technologies.

NBIC Mid Term Review Documentation, March 2011