

<p><b>Institution: University of South Wales</b></p>
<p><b>Unit of Assessment: B10</b></p>
<p><b>Title of case study:</b> Using genomics to shape high performance computing</p>
<p><b>1. Summary of the impact</b></p> <p>Prof. Ron <b>Wiltshire</b> and Dr. Tatiana Tatarinova helped to develop High Performance Computing Wales (HPC Wales) by providing “test” problems from the life sciences (bioinformatics), defining key software packages, and brought on-board collaborative partners to instigate project areas. Established in 2010 with £44 million funding, HPC Wales is Europe’s first nation-wide “hub-and-spoke” super-computing facility. In the three years following its conception, HPC Wales has approximately 50 collaborative organisations supporting over 60 major projects; 40% in life sciences and 5% in bioinformatics through <b>Wiltshire’s</b> and Tatarinova’s involvement. Further impact arose through the redesign of software and up-skilling of HPC Wales staff.</p>
<p><b>2. Underpinning research</b></p> <p><u>The problem</u>  The recent construction of numerous genomic databases has resulted in lengthy lists of millions or billions of chemical units (proteins) that comprise a single organism’s genome. Buried in these lists are the organism’s genes (e.g. it is estimated humans have 20 000-25 000 genes). Searching for known genes within such lists and relating them to their function allows researchers to predict how a particular organism will fare under given conditions (e.g. how a certain person will respond to a particular drug or how a given plant will behave when introduced to a specific soil-type). Hence, targeted and individual-based treatments are possible provided there is sufficient knowledge of the organism’s genetic components. However, searching for these genes within the genome is a computationally intensive task and is the focus of <i>bioinformatics</i>.</p> <p><u>The research</u>  Locating and identifying known genes in an organism’s genome requires a range of parametric and non-parametric statistical approaches to compare data sets, coupled with the development of novel algorithms involving the integration of multiple data types and subsequent efficient processing of those data-rich sets. To this end, significant progress has been made by Dr. Tatiana Tatarinova at the University of South Wales, focussing on partial sequencing of certain plant species with their relatively small sequence length. Tatarinova has published in the region of 40 articles and books and is co-author of over a dozen patents involving the modification of genes in plant crops. Her contribution to the discipline was recognised in the 2010 “International Review of Mathematics in the UK” and she has been rated as one of Wales’s 30 most promising Scientists by Welsh Crucible.</p> <p>Analysing full genome sequences is not feasible using desktop computers due to the sheer scale of the data sets. Instead this requires access to high-performance computing alongside the development of appropriate software exploiting the highly-parallelised architecture of such facilities. Thus in 2010 when HPC Wales was launched, Tatarinova’s research was a natural fit. Indeed, Tatarinova has a history of working for and alongside many of the world’s leading bio-agricultural companies, for example <i>Monsanto Co.</i>, and so has a successful track record in academic and commercial collaborations.</p> <p><u>The impact area</u>  HPC Wales, Europe’s first nationally-distributed supercomputing network, was established in partnership with a number of Welsh universities, the Welsh Government and Fujitsu in 2010. The projects it supports and its long term direction was influenced by its early adopters. Tatarinova’s research in genomics required vast amounts of computational time and was a natural fit with HPC Wales. Therefore Tatarinova, with support from <b>Wiltshire</b>, contributed to the development of HPC Wales by providing important research problems that were used to test and refine the system’s hardware and software. Furthermore, Tatarinova and <b>Wiltshire</b> brought on board a number of collaborative organisations through the promotion of their research specialities. To promote projects in</p>

this area, and as evidence of their support for this work, Fujitsu funded a research studentship while HPC Wales have since funded two further studentships on related projects supervised by Tatarinova.

### 3. References to the research

Research Publications (these publications defined projects and software used in HPC Wales; the de Vere *et al.* paper arose from collaborative work that included the utilization of HPC Wales):

- Tatarinova, T.V.; Alexandrov, N.; Bouck, J.; Feldmann, K. (2010). GC<sub>3</sub> biology in corn, rice, sorghum and other grasses. *BMC Genomics*, **11**: 308. doi: 10.1186/1471-2164-11-308
- Sablok, G.; Chandra Nayak, K.; Vazquez, F.; Tatarinova, T.V. (2011). Synonymous codon usage, GC<sub>3</sub>, and evolutionary patterns across plastomes of three pooid model species: emerging grass genome models for monocots. *Molecular biotechnology*, 49(2): 116-128. doi: 10.1007/s12033-011-9383-9.
- de Vere, N.; Rich, T.C.G.; Ford, C.R.; Trinder, S.A.; Long, C.; Moore, C.W.; Satterthwaite, D.; Davies, H.; Allainguillaume, J.; Ronca, S.; Tatarinova, T.; Garbett, H.; Walker, K.; Wilkinson, M.J. (2012). DNA Barcoding the Native Flowering Plants and Conifers of Wales, *PlosOne*, **7**(6): e37945. doi:10.1371/journal.pone.0037945

Research Grants/Funding (selection of research studentships awarded to Tatarinova that are related to impact on HPC Wales):

- £56K from Fujitsu & HPC Wales for 1 research studentship (No grant number, can provide collaborative agreement on request, but see <http://www.hpcwales.co.uk/farzana-studentship>)
- 2 x KESS Research Students Studentships with Morvus Technology Ltd and National Botanical Gardens Wales, both in collaboration with HPC Wales (Projects “Development of theoretical cell models for the action of novel antiapoptotic proteins” and “Development of bioinformatics tools and procedures for computational support”, respectively)
- £2.5k from Welsh Crucible for “My GENE Code PEiMB: Public Engagement in Molecular Biology”. This Award required selection by the University and attendance at numerous workshops throughout Wales engaging with multi-disciplinary researchers.

### 4. Details of the impact

#### HPC Wales Background

High Performance Computing Wales (HPC Wales), launched in July 2010, is a £44 million five-year project providing a super-computing facility at a scale unprecedented in Europe due to its unique “hub-and-spoke” design. Briefly, a number of “hubs” are distributed throughout the nation and are joined by high-speed connections. Financial support was provided by ERDF and ESF European funds, the Department for Business, Innovation and Skills, collaborating academic institutions (including USW), the Welsh Assembly Government and the private sector.

In early 2011, Fujitsu won the procurement bid to provide infrastructure, support and services. The nature of the projects, the application areas and the necessary equipment was to be shaped by its early adopters since they would be providing real-world problems that would test and refine the system. Associated with the procurement process, Fujitsu re-invested approximately £1.5 million to support projects led by early adopters.

**Wiltshire** identified a number of early adopters, including Tatarinova, who provided sample problems that shaped current application areas, trained HPC Wales staff and introduced further collaborators from commercial organisations, thus helping to expand employment and develop technical skills. Indeed, HPC Wales employs approximately 30 managerial, technical and support staff and it is planned to generate a further 400 jobs requiring high level skills and technical training.

#### Impact on HPC Wales

Tatarinova’s “test” project, started in 2011, involved the analysis of large genome sequences using a combination of standard software coupled with the implementation of novel algorithms. Funded by a £56K award from Fujitsu, HPC Wales were able to understand the type of problems they would be supporting and how to appropriately manage their hardware and software. This work was so influential

## Impact case study (REF3b)

that in September 2012 HPC Wales rated it as one of their top 5 projects.

The success of this project resulted in two further Tatarinova-led projects supported by HPC Wales, each worth £15k, and running in collaboration with non-academic partners including the National Botanical Gardens of Wales (the *Barcode Wales* project) and Morvus Technology Ltd. investigating a range of problems including analysing strains of *Escherichia coli*.

Tatarinova and **Wiltshire** actively promoted HPC Wales by presenting at numerous promotional events and gave media interviews. Further projects and collaborators, both within and outside of academia, arose as a result of these promotional activities (e.g. OSTC Ltd. have two projects utilising HPC Wales working with **Roach**), thus helping to secure the future of HPC Wales and its associated employment.

As of October 2013, and in only three years from its inception, in excess of 50 commercial businesses and industrial partners are engaged in over 60 collaborative projects using HPC Wales; 40% of these involving the life sciences and approximately 5% on problems in genomics and bioinformatics. A key feature of HPC Wales is its accessibility to businesses, such as those brought on board by **Wiltshire** and Tatarinova, and HPC Wales' success in this area was acknowledged through being a runner up in Open Data category of the national Next Generation Digital Challenge Awards in October 2013.

### Impact on supercomputing and bioinformatics

As part of Tatarinova's "test" project, certain bioinformatics software (*cisexpress*) required redesigning to work on the highly-parallelised architecture of HPC Wales. Fujitsu provided the equivalent of approximately £90k in consultancy fees to support the redesign of key algorithms that not only improved essential software for the analysis of large genome sequences, but also resulted in training HPC Wales staff.

Since early 2013, this revised software, running on HPC Wales, has been made available to the world-wide scientific community (<http://glacombio01.comp.glam.ac.uk/cisExpress/new/home.php>) and as of October 2013, approximately 50% of the total simulations performed (numbering several hundred) have been conducted by scientists external to USW. This revised software has therefore laid foundations for further bioinformatics projects around the world with HPC Wales and USW researchers at their core.

### Creating impact legacy

Tatarinova's work with HPC Wales has been designed to create a lasting impact. *Barcode Wales* has constructed a DNA database for the native flowering plants and conifers for Wales and represents the most comprehensive sampling of any national flora to date. Thus, future environmental surveys need not rely on a taxonomic expert to determine flora and can instead use sampled DNA and associated software developed from this project. In January 2013, Tatarinova, alongside Prof. Denis Murphy (USW), created a spin-out company called *myregulome.com ltd* which is currently marketing software for genomic analysis developed in collaboration with Fujitsu and HPC Wales.

## 5. Sources to corroborate the impact

### Shaping HPC Wales:

1. [http://www.hpcwales.co.uk/sites/default/files/hpcwales/HPCWales\\_NewsByte0112.pdf](http://www.hpcwales.co.uk/sites/default/files/hpcwales/HPCWales_NewsByte0112.pdf)  
Details of Fujitsu providing facilities and support for HPC Wales; Tatarinova defining HPC Wales software.
2. <http://www.hpcwales.co.uk/who-we-are>  
Current list of staff employed directly by HPC Wales in management, support and technical roles.
3. <http://www.bbc.co.uk/news/10587005>  
Job creation estimates and funding sources for HPC Wales.
4. <http://www.fujitsu.com/downloads/TC/sc12/booth/hpc-wales-fujitsubp-sc12.pdf>

**Impact case study (REF3b)**

Priority areas of HPC Wales, including a list of customers and customer feedback, demonstrating Tatarinova's influence on shaping HPC Wales. Statement that Tatarinova's "test" project is one of HPC Wales' top 5 projects.

5. <http://www.welshcrucible.org.uk/2011-programme/participants/welsh-crucible-participant-profiles/western-mail-profiles-2011-series/western-mail-profiles-dr-tatiana-tatarinova-university-of-glamorgan/>  
How HPC Wales' projects have been shaped by Tatarinova.
6. [http://www.hpcwales.co.uk/sites/default/files/hpcwales/20120905\\_software\\_list.pdf](http://www.hpcwales.co.uk/sites/default/files/hpcwales/20120905_software_list.pdf)  
Details of software installed on HPC Wales including significant genomics-related programs.

Promoting HPC Wales and introducing business partners:

7. <http://www.bbc.co.uk/programmes/b01pw571>  
Radio interview on BBC Radio Wales on *Science Cafe*, broadcast 15<sup>th</sup> January, 2013, with Tatarinova promoting HPC Wales and explaining how she has helped define its application areas.
8. <http://www.hpcwales.co.uk/fighting-cancer>  
Video link and case study – reference to algorithm being redesigned in collaboration with Fujitsu, HPC Wales and other software developers.

Impact on bioinformatics:

9. International Review of Mathematical Sciences in the United Kingdom  
<http://www.epsrc.ac.uk/SiteCollectionDocuments/other/MathsIR2010EvidenceDocumentsParts1-3.pdf> (p.173)  
"The analysis of properties of genes in various plant species by Tatarinova (Glamorgan) and colleagues has provided major insight for tissue specific and stress response in plants and resulted in patents in areas of plant biotechnology"

Corroborating sources:

10. HPC Wales  
Can corroborate **Wiltshire's** and Tatarinova's involvement in defining software and projects for HPC Wales.