# UCD Impact Case Study

# Clustal: The software at the heart of a biological revolution

Professor Des Higgins UCD School of Medicine

## SUMMARY

In the late 1980s, Professor Des Higgins developed a computer programme – Clustal – that could quickly compare sequences of genetic information. Previously, researchers had to do this arduous task by hand. Although refined over the years, Clustal has always been free to use. It is now a global standard, used hundreds of times a day by scientists, in industry and academia, addressing some of society's biggest challenges. These include controlling diseases, producing biofuels, and creating diseaseresistant plants. As a result, Professor Higgins' Clustal publications are among the world's most highly cited. They are also referenced in thousands of patent documents.

## **RESEARCH DESCRIPTION**

In 1988, Professor Des Higgins developed a computer programme called Clustal. It takes sequences of genetic information (DNA or proteins) and quickly compares them to see what they have in common or how they differ. This is an example of what has become known as bioinformatics – the science of using computers to manage and analyse genetic information.

Professor Higgins wrote the first version of Clustal on a basic IBM computer with very little memory. As word about his software spread, Professor Higgins sent it out on floppy discs to other researchers who were similarly tired of comparing genetic material by hand, using word-processing software.

Over the years, Professor Higgins and colleagues have developed more and more powerful versions of Clustal. But it has always been user-friendly, free to use and able to compare large amounts of genetic information on personal computers. Because of this, the value of the software became apparent to a wide community of scientists, and Clustal became a global standard. Clustal is a global standard, used hundreds of times a day by scientists addressing real-world challenges. Now, one of Professor Higgins' papers on Clustal is among the 10 most highly cited in the world.

ACADEMIC

SCIENTIFIC



Today, the programme is available on most operating systems, and is also delivered as a service over the internet, where it is available on large servers like those at the European Bioinformatics Institute and the Swiss Institute of Bioinformatics.



# **RESEARCH IMPACT**

#### Scientific impact

Thousands of researchers across the globe use Clustal. This ground-breaking work has made Professor Higgins Ireland's most highly cited scientist: his work has been cited in academic papers over 150,000 times. According to the journal Nature, one of Professor Higgins' papers on Clustal is among the 10 most highly cited in the world. He has another in the top 30.

On April 14, 2003, the Human Genome Project was declared complete. This thirteen-year, international project succeeded in working out the DNA sequence of the entire human genome. Since then, the genomes of various other species – including mammals, birds, fish and plants – have also been sequenced, and more are being sequenced every day. Understanding these genomes is crucial for understanding the evolution of life on Earth, and for preventing and curing certain diseases. But genetic information is incredibly complex. It is a significant challenge for biologists understand these genomes, either by comparing them to each other or by comparing different parts of the same genome. Clustal was instrumental in this endeavour.

Recognising these contributions, Bioinformatics.org awarded Professor Higgins the 2018 Benjamin Franklin Award for Open Access in the Life Sciences. The founder and president of Bioninformatics.org said: "The CLUSTAL series are among the very first bioinformatics tools used by any student and are fundamental to the field."

#### Economic and societal impact

Professor Higgins' work has also been cited in thousands of patents. His papers on Clustal are referred to in around 20,000 patent documents, more than 7,000 of which have been issued, for products made by companies like DuPont, Toyota, and Bristol-Myers Squibb. Many of these patents were created to help solve real-world problems, such as modifying plants to resist disease, producing resources like biofuels, and developing new medicines for humans and animals.

Indeed, Clustal is used hundreds of times every day in labs trying to solve some of society's biggest challenges. For example, almost 2,000 publications describe how Clustal was used to increase understanding or control of HIV. Clustal is now routinely used to compare the DNA of different strains of viruses - like HIV or the flu - to track outbreaks of infectious diseases. This is essential for informing policies aimed at minimising the impact of outbreaks.

Other publications describe how Clustal has been used to work out the origins of swine flu, to understand drug resistance in cancer patients, or to deepen our understanding of the evolution of trees. It is practically impossible to track the impacts of all this work on human health, animal health, public policy and the environment. Yet, taken as a whole, it is clear that Professor Higgins' work on Clustal has had profound societal impact. Lives around the world have been transformed by the research that this computer programme has made possible.





# REFERENCES

#### Media coverage

*The top 100 papers*, Nature, 29 October 2014, www.nature.com/news/the-top-100-papers-1.16224 (accessed 7 August 2019)

2018 Benjamin Franklin Award Winner Announced, Bio-IT World, 15 March 2018, www.bio-itworld.com/2018/03/15/2018-benjaminfranklin-award-winner-announced.aspx (accessed 7 August 2019)

#### **Academic references**

Desmond G. Higgins & Paul M. Sharp, *CLUSTAL*: a package for performing multiple sequence alignment on a microcomputer, Gene, Volume 73, Issue 1, 15 December 1988, Pages 237-244, https://doi.org/10.1016/0378-1119(88)90330-7

Julie D. Thompson, Desmond G. Higgins & Toby J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Research, Volume 22, Issue 22, 11 November 1994, Pages 4673-4680, https://doi.org/10.1093/nar/22.22.4673

Julie D. Thompson, Toby J. Gibson, Frédéric Plewniak, François Jeanmougin & Desmond G. Higgins, The CLUSTAL\_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools, Nucleic Acids Research, Volume 25, Issue 24, 1 December 1997, Pages 4876–4882, https://doi.org/10.1093/nar/25.24.4876

#### **Clustal services**

Clustal Omega, hosted by the European Bioinformatics Institute, https://www.ebi.ac.uk/Tools/msa/clustalo/ (Accessed 7 August 2019)

Clustal W, hosted by the Swiss Institute of Bioinformatics, https://embnet.vital-it.ch/software/ClustalW.html (Accessed 7 August 2019)

#### **Patents and citations**

Google Patent search results for "Clustal", https://patents.google.com/?q=Clustal, (19,533 results, 7,303 when filtered for granted patents only, accessed 7 August 2019)

Google Scholar profile for Professor Des Higgins, https://scholar.google.co.uk/citations?user=ApOK7rUAAAAJ (173,220 citations, accessed 7 August 2019)