

# Daniel MacLean

Born: September 8, 1978

Nationality: British

email: [macleandaniel@gmail.com](mailto:macleandaniel@gmail.com)

url: <http://www.danmacleandaniel.info>

---

## Appointments held

- 2010-Present *Head of Bioinformatics*, The Sainsbury Laboratory, Norwich, UK
- 2010-Present *Honorary Lecturer*, School of Computational Sciences, University of East Anglia, Norwich UK
- 2009-Present *Head of Training in Bioinformatics*, The Sainsbury Laboratory, Norwich, UK
- 2006-2010 *Bioinformatician*, The Sainsbury Laboratory, Norwich, UK
- 2005-2006 *Post-Doctoral Researcher*, Carnegie Institution Department of Plant Biology, Stanford, CA, USA

## Responsibilities

The Sainsbury Laboratory, Norwich, UK

### Head of Bioinformatics

My role is to provide the research labs with primary bioinformatics support and advise on appropriate techniques and applications for achieving project research goals. My areas of expertise include genomics, in particular assembly and alignment with high-throughput sequencing data, genetic polymorphism detection, homology search and automatic sequence annotation, transcriptomics (microarray and RNA-seq), statistical analysis, mathematical and systems-biology approaches to modelling and visualising data and large scale phosphoproteomics analysis.

I've implemented a novel biologist-first bottom-up support model that prioritises the needs of the scientist and makes support provision flexible and scalable in the Big Data era. The success of our bioinformatics support model has been such that all the scientists of TSL groups with projects with informatics components are themselves routinely able to carry out complex bioinformatics led experiments without direct intervention from a dedicated bioinformatician. Our model aims to bring tool use and understanding into the same heads as the biology knowledge to maximise research insights.

I have developed a range of tools inspired by the needs of the research groups at The Sainsbury Laboratory, including the creation of tools for the delivery and visualisation of next-generation sequencing data, contributed to the Bio-Ruby project with a SAMtools gem, tools for the detection of phosphorylated sites in peptides and identification of sRNA loci in genomes from next-generation sequencing data. I also maintain a toolkit of modules and scripts that can be used by all groups across the laboratory. All these tools are in constant development and can be

seen on my github page <http://github.com/danmaclean>. I've developed web-service systems for interrogation and visualisation of next-generation sequencing data. I hold a BBSRC grant funding a project in which we are using Next-Gen sequencing data to identify disease-resistance linked heterozygous mutations without a reference sequence. Recently, I have modelled the *hrpL* regulon of *Pseudomonas syringae* in a Boolean framework, created and implemented algorithms for detection of sRNA loci from high-throughput sequencing data and analysed complex networks of sRNAs in *Arabidopsis*.

I have a key role in the hiring process for the current members of our bioinformatics team, both in the core group and the individual research groups. Day-to-day I line-manage the activities of the bioinformatics assistant and a Post-Doctoral scientist working on a project of my devising and funded from Sainsbury Lab core grants. I recently led TSL's successful application to EMBO for funds supporting a 12 day practical course at TSL in 2012.

As the person responsible for directing the bioinformatics service provision I work very closely with the IT team to ensure that the systems and infrastructure provided is fit for purpose and accessible to users.

#### Head of Training in Bioinformatics

In this role I am responsible for ensuring that Sainsbury Lab scientists have the bioinformatic skills they require to successfully carry out their analyses. I manage the training budget to provide materials and resources for in-house and external courses. These include beginner's courses, tutorials in specific issues and master classes designed to bring together the expertise growing in the lab. I administer a website that collates and distributes these materials at <http://tsltraining.tsl.ac.uk>.

#### Honorary Lecturer, School of Computational Sciences, UEA

I teach and examine students on the Computational Masters degree course. Each year I supervise at least one student to work on projects that become the subject of the dissertation on which their degree is examined. I have trained a total of five Masters students to date. I am also a member of the advisory board for this course, a body that is responsible for the direction and syllabus of the course.

#### Prior Research Experience

2005 – 2006 Carnegie Institution, Dept. of Plant Biology, Stanford University, CA, USA  
Post-doctoral researcher, laboratory of Dr Sue Rhee

I used informatics approaches to examine system levels patterns of transcriptional response to multiple abiotic stresses in *Arabidopsis* from public data and create a holistic model of the stress response system. Orthologue detection was used to assess the evolutionary derivation of sub-systems. I used Bayesian modeling systems, machine learning methods and promoter analysis to predict transcriptional interaction networks and create models of stress regulatory networks. Phylogenetic analysis was used to model potential evolution of transcription factor families involved in stress response.

2001- 2005 Department of Plant Sciences, University of Cambridge, UK  
Ph.D, supervised by Professor John C. Gray  
Transcriptional Regulation of Nuclear Genes for Plastid Proteins I examined transcriptional regulation of nuclear genes for plastid proteins. Custom and Affymetrix microarrays were used. I developed basic software to analyse these custom arrays. I applied bioinformatics methods to create models of transcriptional networks for observed transcriptional co-regulation. Potential regulatory regions for nuclear genes were identified and promoter-reporter fusions were used to assess their role.

### Education

2005 Ph.D ‘Transcriptional regulation of nuclear genes of the plastid genetic system’, Darwin College, University of Cambridge  
2001 M.Res Biological Sciences, University of Manchester  
2000 B.Sc (Hons) Biological Sciences, University of Manchester

### Expertise

Genomics: *de-novo* assembly, re-sequencing, mapping, gene prediction, comparative genomics, SNP analysis

Transcriptomics: RNA-seq, microarray analysis

Statistics and Mathematics: Parametric and non-parametric statistical methods, graph/network analysis

Informatics: Scripting, web development, algorithm design and implementation

### Technical Skills

#### **Operating systems**

Experienced: Linux (Debian and derivatives), Mac OS X, Windows XP

#### **Programming**

Experienced: Perl (including BioPerl and Object Oriented Perl), Ruby, BioRuby, BioPython (including NetworkX and BooleanNet), Javascript, HTML, CSS, MySQL

#### **Software and Frameworks**

Experienced: R, BioConductor, L<sup>A</sup>T<sub>E</sub>X, BLAST, BLAT, ssaha, Velvet, maq, Bowtie, Cortex, Novoalign, Oases, BWA, Rails, Gbrowse, EMBOSS, SAMtools

Familiar: PFAMScan, miRanda, mirCat, CLUSTALW, phylip, InParaNoid, Hmmer, libsvm, SciLab

## Personal Interests

Blogging photographer, both traditional print and digital media <http://www.onandoff.co.uk>

Keen guitarist

## Grants & awards

BBSRC Tools And Resources Fund 2011

EMBO Practical Course Funding to TSL for Plant-Pathogen Interactions 2012

Roberts Funding for training, regarding training of Sainsbury Laboratory staff in Bioinformatics

Keystone Symposia Scholarship meeting on Plant Responses to Abiotic Stress 2006

Company of Biologists Travel Grant 2004

Darwin College Travel Fund 2002 and 2004

Frank Smart Studentship 2002 and 2003

## Memberships and professional activities

Invited Member IAIC Design Workshop, Atlanta, USA, 2011 Member MSc Computational Biology Advisory Board University of East Anglia, 2009 - Present Member ISCB 2009 - Present Reviewer/Editorial Board Member - Nature Biotechnology, Bioinformatics, BMC Bioinformatics, Current Bioinformatics

## Publications & talks

Times Cited - 642, Average Citations per Item - 22.14, h-index - 13

### Journal articles

2011 MacLean D and Kamoun S "Squeezing Big Data into a Small Organisation" *Nat Biotech in press*

Richard M. Leggett, Ricardo Ramirez-Gonzalez, Walter Verweij, Zamin Iqbal, Jonathan D. G. Jones, Mario Caccamo, Daniel MacLean Bubbleparse: a tool for identifying SNPs directly from NGS reads, without need for a reference genome. *Under review at Nat Biotech*

Ramirez-Gonzalez R, Caccamo M, MacLean D. Gee Fu: a sequence version and web-services database tool for genomic assembly, genome feature and NGS data (2011) *Bioinformatics* 27, 2754-2755

Earl D et al. Assemblathon 1: A competitive assessment of *de novo* short read assembly methods. (2011). *Genome Res.* 21, 2224-2241

Eric Kemen, Anastasia Gardiner, Torsten Schultz-Larsen, Ariane C. Kemen, Alexi L. Balmuth, Alexandre Robert-Seilaniantz, Kate Bailey, Eric Holub, David J. Studholme, Dan MacLean,

Jonathan D. G. Jones. (2011) Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis*. *PLoS Biol* 9, e1001094

Robert-Seilaniantz A, Maclean D, Jikumaru Y, Hill L, Yamaguchi S, Kamiya Y, Jones JD. (2011). The microRNA miR393 redirects secondary metabolite biosynthesis away from camalexin and towards glucosinolates. *Plant J.* 67, 218-231

Ruby Biogems: pluggable software modules for the next generation. Raoul J.P. Bonnal et al  
*Under review at Bioinformatics*

Etherington G and MacLean D (2011) Next-Generation techniques for analysing Next-Generation sequence data (2011). *GIT Laboratory Journal*, 15, 11-12, pp 10-11

Ricardo Ramirez-Gonzalez, Raoul JP Bonnal, Mario Caccamo, Daniel MacLean samtools-ruby: a Ruby interface for the SAMtools library. *Under Review at Open Research Computation*

2010 MacLean D, Yelina N, Havecker E, Heimstadt S, Studholme DJ and Baulcombe DC (2010) Evidence for Large Complex Networks of Plant Short Silencing RNAs. *PLoS ONE*, 5, e9901

MacLean D and Studholme DJ (2010) A Boolean model of the *Pseudomonas syringae* hrp regulon predicts a tightly regulated system *PLoS ONE*, 5, e9101

MacLean D, Moulton V and Studholme DJ Identification of small RNA generative loci from high-throughput sequencing data using NiBLS. *BMC Bioinformatics* 11, 93

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RH, Zody MC, Kunjeti SG, Donofrio NM, Meyers BC, Nusbaum C, Kamoun S. (2010) Genome Evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330, 1540-1543

Lewsey MG, Murphy AM, MacLean D, Dalchau N, Westwood JH, Macaulay K, Bennett MH, Moulin M, Hanke DE, Powell G, Smith AG and Carr JP. (2010) Disruption of two defensive signalling pathways by a viral RNA silencing suppressor. *Mol Plant Microbe Interact.* 23:835-845.

Studholme DJ, Kemen E, MacLean D, Schornack S, Aritua V, Thwaites R, Grant M, Smith J, Jones JD. (2010) Genome-wide sequencing data reveals virulence factors implicated in banana Xanthomonas wilt. *FEMS Microbiol Lett.* 310, 182-192.

Durek P, Schmidt R, Heazlewood JL, Jones A, MacLean D, Nagel A, Kersten B and Schulze WX. (2010) PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Res.* 38, 828-384.

Salomon S, Grunewald D, Stüber K, Schaaf S, MacLean D, Schulze-Lefert P and Robatzek S. (2010) High throughput confocal imaging of intact live tissue enables quantification of membrane trafficking in *Arabidopsis*. *Plant Phys.* in press

- Comparot-Moss S, Kötting O, Stettler M, Edner C, Graf A, Weise SE, Streb S, Lue WL, MacLean D, Mahlow S, Ritte G, Steup M, Chen J, Zeeman SC, Smith AM. (2010) A putative phosphatase, LSF1, Is Required for normal starch turnover in *Arabidopsis* Leaves. *Plant Physiol.*, 152, 685-697.
- 2009 MacLean D, Jones JD and Studholme D (2009) Application of 'Next Generation' sequencing technologies to microbial genetics, *Nature Reviews Microbiology*, 7, 287-296
- Jones AM, MacLean D, Studholme DJ, Serna-Sanz A, Andreasson E and Peck SC (2009) Phosphoproteomic analysis of nuclei-enriched fractions from *Arabidopsis thaliana* *Journal of Proteomics*, 72, 439-451
- Studholme DJ, Ibanez SG, MacLean D, Dangl JL, Chang JH and Rathjen JP. A draft genome sequence and functional screen reveals the repertoire of type III secreted proteins of *Pseudomonas syringae* pathovar tabaci 11528. *BMC Genomics* 24, 395
- Haas BJ, Kamoun S *et al* (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461, 393-398
- Almeida NF, Yan S, Lindeberg M *et al* (2009) A draft genome sequence of *Pseudomonas syringae* pv. tomato strain T1 reveals a repertoire of type III related genes significantly divergent from that of *Pseudomonas syringae* pv tomato strain DC3000, *MPMI* 22, 52-62
- Valkov VT, Scotti N, Kahlau S, MacLean D, Grillo MS, Gray JC, Bock R and Cardi T (2009) Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: transcriptional and post-transcriptional control, *Plant Phys.* 150, 2030-2044
- 2008 MacLean D, Burrell MA, Studholme DJ and Jones AMEJ (2008) PhosCalc: A tool for determining the likelihood of peptide phosphorylation from Mass Spectrometer data. *BMC Research Notes* 1:30
- MacLean D, Jerome CA, Brown AP, Gray JC (2008) Co-regulation of nuclear genes encoding plastid ribosomal proteins by light and plastid signals during seedling development in tobacco and *Arabidopsis*. *Plant Mol Biol.* 66, 475-490
- Moxon S, Schwach F, Dalmay T, MacLean D, Studholme DJ, Moulton V (2008) A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 24, 2252-2253
- van den Burg HA, Tsitsigiannis DI, Rowland O, Lo J, Rallapalli G, MacLean D, Takken FL, Jones JD (2008) The F-Box Protein ACRE189/ACIF1 Regulates Cell Death and Defense Responses Activated during Pathogen Recognition in Tobacco and Tomato. *Plant Cell.* 20, 697-719
- 2007 Cottage AJ, Mott AK, Wang JH, Sullivan JA, MacLean D, Tran L, Choy MK, Newell CA, Kavanagh TA, Aspinall S, Gray JC (2007) GUN1 (GENOMES UNCOUPLED1) encodes a pentatricopeptide repeat (PPR) protein involved in plastid protein synthesis-responsive retro-

grade signaling to the nucleus. In: Allen J, Gantt E, Golbeck J, Osmond B (eds) *Energy from the Sun, Proceedings of the International Photosynthesis Congress. Springer, Berlin*

Anastasiou E, Kenz S, Gerstung M, MacLean D, Timmer J, Fleck C, Lenhard M (2007) Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. *Dev Cell*. 13, 843-856

Magee AM, MacLean D, Gray JC, Kavanagh TA. (2007) Disruption of essential plastid gene expression caused by T7 RNA polymerase-mediated transcription of plastid transgenes during early seedling development. *Transgenic Res.* 4, 415-428

2004 Chua YL, Mott E, Brown AP, MacLean D, Gray JC (2004) Microarray analysis of chromatin-immunoprecipitated DNA identifies specific regions of tobacco genes associated with acetylated histones. *Plant J.* 37, 789-800.

2003 Gray JC, Sullivan JA, Wang J-H, Jerome CA and MacLean D (2003) Coordination of plastid and nuclear gene expression. *Phil Trans R Soc Lond B* 358, 135-145

MacLean D. Report on the Transcriptomics Session of the 2003 Main Meeting of the SEB. *SEB Bulletin*, July 2003.

#### Book Chapters

2011 Walshaw J, Etherington G and MacLean D (2011) Next-Generation Sequencing Approaches to Metagenomics In D Marco (ed) *Metagenomics*, Caister Academic Press, UK

Daniel MacLean “Big data in a small place. A case study in scalable lightweight core-support” Invited book chapter “*Free and Open Source Software in Applied Life Science and Industry: Practical Solutions to Common Problems*” Eds Lee Harland and Mark Forster, BioHealthcare Publishing.

2009 MacLean D and Studholme DJ (2009) Bioinformatics aspects of high-throughput sequencing. In RW Jackson (ed), *Plant Pathogenic Bacteria: Genomics and Molecular Biology*, Caister Academic Press, UK

#### Invited talks and presentations

2011 MacLean D, Big Data in Small Places, Bio-IT World, Hannover, Germany, 2011  
MacLean D, Squeezing Big Data into a Small Organization, Eagle Genomics Symposium, Babraham, UK, 2011

2009 MacLean D, Finding Resistance-Linked SNPs in Potato with Illumina Data. ICRISAT, India, 2009

2006 MacLean D, Keystone Symposium on Plant Responses to Abiotic Stress. System Level Response to Abiotic Stresses in Arabidopsis Keystone Symposia, Copper Mountain, CO, USA, 2006

2003 MacLean D, Society for Experimental Biology Meeting Examining chloroplast genome gene

expression with a custom microarray. Transcriptomics Session, Southampton, UK, 2003

2002 MacLean D, Plastid Preview 2002. Microarray analysis of plastid transcripts. Leicester, UK, 2002

#### Posters

2010 MacLean D gee\_fu: an extensible Ruby-on-Rails database and web-service application for delivering HTGS data *VizBi: Visualising Biological Information*, EMBL, Heidelberg, Germany, 2010

2009 MacLean D and Jones AME PhosCalc: A tool for evaluating the sites of peptide phosphorylation from Mass Spectrometer data. *Proteomics Day*, John Innes Centre, UK, 2009

2008 MacLean D, Yelina N, Havecker E, Heimstadt S, Studholme DJ and Baulcombe DC sRNAs regulate each other in large complex networks. *International Conference on Systems Biology*, Gothenburg, Sweden, 2008.

Kemen E, Studholme DJ, MacLean D and Jones DG, Identification of Novel Effectors from *Albugo candida* using Solexa sequencing technologies. *Keystone Symposia*, Keystone, CO, USA, 2008.

Lewsey M, MacLean D and Carr J RNA silencing in SA induced defence. *Keystone Symposia*, Keystone, CO, USA, 2008.

2006 MacLean D and Rhee SY System Level Response to Abiotic Stresses in *Arabidopsis* *Keystone Symposia*, Copper Mountain, CO, USA, 2006.