DNA sequencing tackles global challenges

BBSRC bioscience for the future

When Watson and Crick first described the structure of DNA in 1953, they laid the foundations of modern bioscience. Subsequent discoveries enabled researchers to determine the sequence of the four 'bases' that form the building blocks of the helical DNA.

Knowing an organism's DNA sequence enables researchers, often supported by BBSRC investments, to understand how it functions. For instance, understanding which genes a plant or animal carries and where those genes are in the genome allows breeders to select for beneficial variants in their breeding programmes, leading to livestock and crops with higher yield, which are resistant to disease, or which thrive in specific environments. Alternatively, knowing the genetic basis of health can help clinicians develop and target treatments for illnesses with a genetic basis. DNA sequencing can also help researchers identify potentially valuable natural products and produce new and improved strains of microbes used to manufacture biological products such as medicines or industrial enzymes.

BBSRC investments in the early 1990s supported some of the earliest published DNA sequences (yeast and the model plant *Arabidopsis*, used in research globally). BBSRC also supported the research that underpinned spinout company Solexa. Solexa, subsequently bought by Illumina for $600M, developed one of the most widely-used sequencing technologies, which was much faster and cheaper than previous methods. More recently, BBSRC-funded research underpinned the formation of spinout Oxford Nanopore to develop and commercialise their innovative new sequencing technology.

BBSRC-funded researchers have led and contributed to international consortia sequencing commercially- and societally-important species including bread wheat, cows, diseases such as ash dieback and the chicken pathogen *Eimeria*. Researchers with funding from BBSRC and others are also contributing to major national initiatives such as the UK Government's 100,000 Genomes project, which aims to understand the genetic basis of cancers and other rare diseases.

This timeline provides a snapshot of some of the major BBSRC investments in DNA sequencing and its application to address some of the most important challenges facing society today.

BBSRC support for sequencing technology

1996
BBSRC funding of £282K awarded to Dr Shankar Balasubramanian and Dr David Klenerman at the University of Cambridge, to support their working developing a new approach to DNA sequencing.[1]

1998
Spinout company Solexa founded by Balasubramanian and Klenerman to commercialise a technique to sequence DNA based on BBSRC-funded research.[1]

2000
ARK Genomics is set up by BBSRC as part of its 'Investigating Gene Function' initiative to provide access to functional genomics technology and resources relevant to the animal health community.

2004
Human Genome project completed by US researchers, revealing more than 20,000 genes. The project took 13 years and cost around $2.7Bn (in 1991 US dollars).[2]

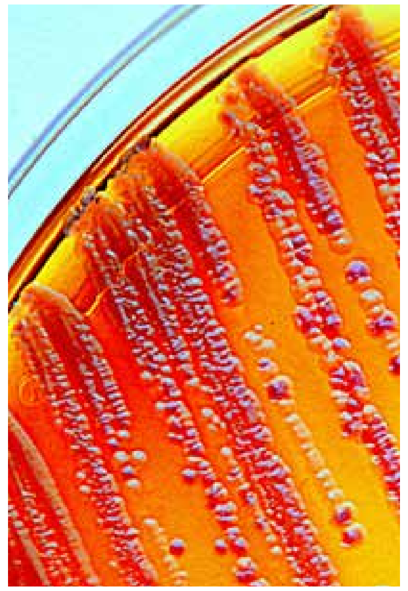
2005
Oxford Nanolabs Ltd (later renamed Oxford Nanopore Technologies) founded as a spinout company by Oxford University, based on research from BBSRC-funded studentships.[3]

2006
Solexa launches its first sequencer – the Genome Analyzer – capable of sequencing one gigabase of data in a single run.
In the same year, Solexa acquires Lynx Therapeutics (an instrumentation company) and becomes an international public company.[4]

2007
US genetics analysis company Illumina buys Solexa for $600M.[5]

2009
New UK national research institute launched; The Genome Analysis Centre (TGAC). TGAC receives strategic funding from BBSRC and provides core national capabilities in genome sequencing and annotation.[6]

2010
Professor Balasubramanian named BBSRC Innovator of the Year for inventing Solexa sequencing and for his work launching Solexa Ltd and the subsequent commercialisation and dissemination of the technology.[7]



2012
Oxford Nanopore Technologies announces new hand-held sequencing technology MinION and GridION. BBSRC training grants helped to prove that biological nanopores could differentiate individual bases in a single DNA molecule – a key aspect of the ONT sequencing method.[8]

2012
Advances in sequencing technology mean that the human genome can now be sequenced in one day.

2013
ARK Genomics receives a £1.1M funding boost from BBSRC to support its new status as a National Capability, providing access to next-generation sequencing, high-throughput genotyping, microarrays and bioinformatics.[9] It is subsequently merged with The GenePool at Edinburgh University to create Edinburgh Genomics.

The international NORNEX consortium, led by JIC, is awarded BBSRC funding to investigate the molecular and cellular basis of interactions between the Ash Dieback fungus and ash trees. The sequence of a survivor tree is completed the same year and is made available for analysis on a crowdsourcing site.[10]

Elixir, a pan-European research infrastructure for biological information such as sequencing data, is formally launched. Elixir aims to orchestrate the collation, quality control and archiving of biological data; the UK node focuses on training. BBSRC is one of the UK sponsors.[11]

2014
ONT launches the MinION USB device for early access and testing – TGAC is one of the recipients of the early devices. BBSRC (and NERC) funded software to analyse the data produced is developed by Edinburgh Genomics.[12] In the same year, ONT is valued at $2Bn.[13]

2014
Illumina claims around 70 % of the $2.5Bn global sequencing market.[14]

2014
Cost of sequencing the entire human genome falls to $1000 using Illumina technology.[15]

2015
ONT charge just $1000 for researchers to access their MinION sequencing technology.[16]

1990 | 2000 | 2005 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015

1992
First complete chromosome sequence; chromosome III from yeast (*Saccharomyces cerevisiae*).[17]

1996
The first full genome sequence of a eukaryote; the yeast *Saccharomyces cerevisiae*.[18]

2000
First plant genome to be sequenced; the model plant *Arabidopsis thaliana*, which is used as a model plant in research globally. The Arabidopsis Genome Initiative was part-funded by BBSRC.[19]

2002
Streptomyces coelicolor genome published, led by researchers at the John Innes Centre (JIC), which receives strategic funding from BBSRC. Streptomyces bacteria produce a wide range of potentially valuable 'specialised metabolites', including new antibiotics.[20]

2004
Chicken (*Gallus gallus*) genome published. Researchers funded by BBSRC contributed to the sequencing and analysis, the establishment of a library of short DNA sequences that span the entire genome, and the development of tools to help probe the function of individual genes.[21]

2005
Rice (*Oryza sativa*) map-based sequence published. The international consortium included researchers at JIC.[22]

2007
Chicken pathogen *Eimeria tenella* sequenced, supported in part by BBSRC. Eimeria infections cause coccidiosis, which costs the global poultry industry £2Bn per year.[23]

2009
Cattle (*Bos taurus*) genome published. UK scientists, supported by BBSRC, played a key role in the annotation and analysis of the genome as part of a 300-scientist international collaboration.[24]

2010
Pea aphid (*Acyrthosiphon pisum*) sequenced. Researchers at Rothamsted Research, which receives strategic funding from BBSRC, were among those involved.[25]

Grass species purple false brome (*Brachypodium distachyon*) sequenced. Researchers from JIC were among those involved.[26]

First sequence coverage of the wheat (*Triticum aestivum*) genome is published by UK researchers with funding from BBSRC.[27]

Strawberry (*Fragaria vesca*) sequenced. BBSRC contributed to the work of UK scientists involved in this international consortium.[28]

Barley Powdery mildew (*Blumeria graminis f.sp. hordei*) sequenced. BBSRC contributed to the international effort to sequence this important barley pathogen.[29]

2011
Potato (*Solanum tuberosum*) genome published. The UK component of the potato genome sequencing consortium was part-funded by BBSRC grants to the University of Dundee and Imperial College London.[30]

Model legume *Medicago truncatula* sequenced. BBSRC supported the work at the Wellcome Trust Sanger Institute and JIC.[31]

Septoria wheat blotch (*Mycosphaerella graminocola*) sequenced. This international project involved researchers at Rothamsted.[32]

Chinese cabbage (*Brassica rapa*) sequenced; the project involved researchers at JIC and Rothamsted.[33]

2012
Wheat genome sequence annotation published. BBSRC contributed to this project by supporting researchers at Bristol, Liverpool and JIC.[34]

Pig (*Sus scrofa*) genome published. The UK component of the international swine genome sequencing consortium was part-funded by BBSRC.[35]

Tomato (*Solanum lycopersicum*) genome published. UK researchers, focussing on chromosome four, set the standard for the quality of sequence produced internationally. The UK team were part-funded by BBSRC.[36]

High resolution draft barley (*Hordeum vulgare L.*) genome sequence published. The UK component (involving James Hutton Institute and TGAC) of the international barley genome sequencing consortium was part-funded by BBSRC.[37]

2013
Sequence of the marine alga *Emiliana Huxley*. This alga produces omega-3 long chain polyunsaturates, and is important in fixing atmospheric carbon. This international project involved Rothamsted.[38]

BBSRC researchers sequence an ash tree (*Fraxinus excelsior*) resistant to the ash dieback fungus (*Hymenoscyphus fraxineus*).[39]

2014
Twenty genomes of the ash dieback fungus are sequenced by TGAC and added to the crowdsourcing site OpenAshDieBack.[40]

Ferret (*Mustela putorius furo*) genome sequenced by an international team that included researchers from TGAC. Ferrets are used as models of certain human diseases, such as influenza.[41]

First whole-genome draft of the annotated wheat genome sequence released by the International Wheat Genome Sequencing Consortium. BBSRC contributed grant funding to this international project and the sequencing was performed at TGAC.[42]

TGAC contributes to sequencing the agriculturally-important nitrogen-fixing soil bacterium *Azospirillium brasilense*.[43]

TGAC involved in sequencing the naked mole rat (*Heterocephalus glaber*). This model organism is long-lived and is resistant to developing cancers; it is therefore used in research into longevity and cancer.[44]

2015
Sequencing of the first 48 yeast strains from the BBSRC-supported National Collection of Yeast Cultures (maintained at the Institute of Food Research, which receives strategic funding from BBSRC) is completed by TGAC. TGAC also plan to sequence the full collection of approximately 4000 strains.[45]

Red clover (*Trifolium pratense L.*) genome sequenced by researchers at TGAC and IBERS. Red clover is used as a protein-rich livestock feed, and can also help enrich soils with nitrogen.[46]

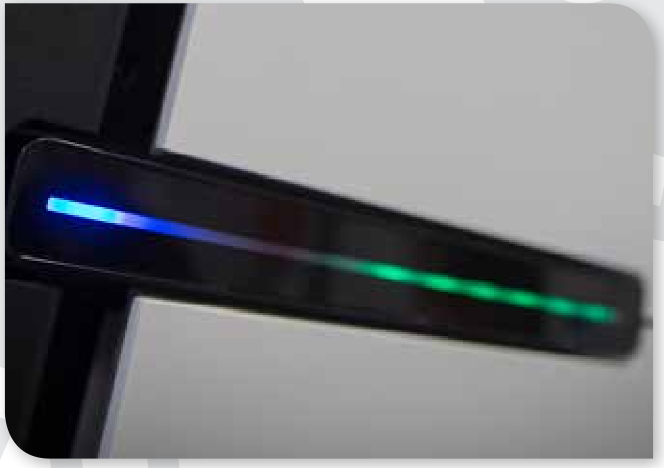
Examples of genome sequencing supported by BBSRC

Solexa and Illumina

Based on BBSRC-funded research in the 1990s, Professor Balasubramanian and colleagues from the University of Cambridge developed 'Solexa' sequencing technology. Using Solexa, researchers could sequence DNA 10,000 faster and much more cost-effectively than was previously possible. In 1998 Professor Balasubramanian and Professor Klenerman founded Solexa Ltd to develop and commercialise the technology. The company was later bought by Illumina for US$600M. According to Illumina, today more than 90% of the world's sequencing data is generated using their sequencing by synthesis (SBS) technology.[47] Professor Balasubramanian was named as the 2010 BBSRC Innovator of the Year for inventing Solexa sequencing, and for his work in launching Solexa Ltd, and the subsequent commercialisation and dissemination of the technology.

Advances using Illumina sequencing technology:

- Illumina sequencing was used by an international consortium of researchers to sequence the genome of wheat; a staple crop providing 20% of the world's calories.[48] Using the wheat genome sequence, breeders are beginning to understand the genetics of complex characteristics such as yield stability and resistance to pests and disease.[49] Detailed mapping of the location of these genes is helping plant breeders to develop improved wheat varieties more quickly and efficiently.



- Scientists are studying the genome sequences of the chicken and the chicken pathogen Eimeria (both sequenced using Illumina technology) to improve welfare and efficiency in chicken production by reducing disease in farmed poultry. The Eimeria sequence is being used to identify potential new targets for vaccines or drugs against coccidiosis, a devastating disease of chickens which costs farmers £2Bn per year[50], contributing to global food security. The chicken sequence has also been used to breed chickens that do not spread bird flu to other chickens. It is also being used to study the avian immune system, with the aim of reducing the quantity of chemicals and antibiotics that are used in poultry production.

- Sequencing is being used to help in the battle against the ash dieback epidemic. Sequencing the fungus responsible for the disease is increasing our understanding of how it infects and is spread between trees, leading the way towards development of effective treatments. Researchers have also sequenced ash trees with naturally high levels of resistance to ash dieback to identify the genes responsible for their resistance. This could help breeders create resistant strains of ash tree to replace those lost to the disease.

Sequencing in the news

In 2012, BBC news reported that researchers had used Illumina sequencing to stop an outbreak of MRSA in a hospital.

By analysing the genome sequences of MRSA isolates form a special baby care unit at Rosie Hospital in Cambridge, researchers traced them all to a single source of infection. This source was identified as a member of staff who was a carrier of MRSA but showed no symptoms; once treated further MRSA infections in the outbreak stopped.

- The UK's 100,000 Genomes Project, was established in 2014 to sequence the genomes of NHS patients with rare diseases and cancer, as well as their families. Illumina is working with the project to deliver the necessary infrastructure and expertise.[51] The project is the largest such national project in the world and is being managed by Genomics England. It was set up by the UK Government with the aims of establishing a 'genomics medicine' service for the NHS, enabling medical research, and building a UK genomics industry. Data analysis for the project is being supported by spinout company Genomics plc, which arose from BBSRC-funded research at the University of Oxford to develop new DNA sequence analysis software.

BBSRC bioscience for the future

References

1 www.bbsrc.ac.uk/news/people-skills-training/2014/140728-f-gb-bioscience-pioneers-shankar-balasubramanian
2 International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-45. doi: 10.1038/nature03001. Available online at www.nature.com/nature/journal/v431/n7011/full/nature03001.html. And: www.genome.gov/11006943
3 [Reference/webpage no longer available – July 2016]
4 www.illumina.com/technology/next-generation-sequencing/solexa-technology.html
5 www.illumina.com/technology/next-generation-sequencing/solexa-technology.html
6 [Reference/webpage no longer available – July 2016]
7 www.bbsrc.ac.uk/research/impact/solexa-sequencing-innovator-winner
8 [Reference/webpage no longer available – July 2016]
9 [Reference/webpage no longer available – January 2017]
10 [Reference/webpage no longer available – July 2016]
11 www.elixir-europe.org/about/elixir-structure
12 TGAC takes part in MinION Access Programme (MAP) for portable DNA sequencing. TGAC-takes-part-in-MinION-Access-Programme-MAP-for-portable-DNA-sequencing/; Watson M, Thomson M, Risse J, Talbot R, Santoyo-Lopez J, Gharbi K, Blaxter M (2014). poRe: an R package for the visualization and analysis of nanopore sequencing data. *Bioinformatics* 13(1):114-5. doi: 10.1093/bioinformatics/btu590
13 REF case study: 'Oxford Nanopore Technologies: a successful company built on innovative DNA sequencing', University of Oxford. Available online at: http://results.ref.ac.uk/DownloadFile/ImpactCaseStudy/pdf?caseStudyId=18129
14 www.marketsandmarkets.com/PressReleases/ngs-technologies.asp
15 Science enters $1000 genome era, www.bbc.co.uk/news/science-environment-25751958
16 [Reference/webpage no longer available – July 2016]
17 Oliver S, et al (2992). The complete DNA sequence of yeast chromosome III. *Nature* 357:38-46. doi:10.1038/357038a0. Available online at www.nature.com/nature/journal/v357/n6373/abs/357038a0.html
18 Goffeau A et al, (1996). Life with 6000 genes. *Science* 3274:546-67. Doi: 10.1126/science.274.5287.546. Available online at www.sciencemag.org/content/274/5287/546
19 The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815. doi: 10.1038/35048692. Available online at www.nature.com/nature/journal/v408/n6814/full/408796a0.html
20 Bentley S, et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417: 141-7. doi:10.1038/417141a. Available online at www.nature.com/nature/journal/v417/n6885/full/417141a.html
21 International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018):695-716. doi: doi:10.1038/nature03154. Available online at: www.nature.com/nature/journal/v432/n7018/full/nature03154.html
22 International Rice Genoma Sequencing Project (2005). The map-based sequence of the rice genome. *Nature* 436:793-800. doi:10.1038/nature03895. Available online at www.nature.com/nature/journal/v436/n7052/full/nature03895.html
23 Ling KH, Rajandream MA, Rivailler P, Ivens A, Yap SJ, Madeira AM, Mungall K, Billington K, Yee WY, Bankier AT, Carroll F, Durham AM, Peters N, Loo SS, Isa MN, Novaes J, Quail M, Rosli R, Nor Shamsudin M, Sobreira TJ, Tivey AR, Wai SF, White S, Wu X, Kerhornou A, Blake D, Mohamed R, Shirley M, Gruber A, Berriman M, Tomley F, Dear PH, Wan KL.(2007) Sequencing and analysis of chromosome 1 of *Eimeria tenella* reveals a unique segmental organization. *Genome Res*. 17(3):311-9.
24 The Bovine Genome Sequencing and Analysis Consortium (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324: 522-8. doi: 10.1126/science.1169588. Available online at http://science.sciencemag.org/content/324/5926/522.full
25 The International Aphid Genomics Consortium (2010). Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *PLoS Biol* 8(2): e1000313. doi:10.1371/journal.pbio.1000313. Available online at http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000313
26 The Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463:763-8. doi: 10.1038/nature08747. Available online at www.nature.com/nature/journal/v463/n7282/full/nature08747.html
27 Press release: 'UK researchers release draft sequence coverage of wheat genome', University of Bristol. Available online at www.bristol.ac.uk/news/2010/7186.html
28 Shulaev V, et al (2010). The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43:109–116. doi:10.1038/ng.740. Available online at www.nature.com/ng/journal/v43/n2/full/ng.740.html
29 Spanu P, et al (2010). Genome Expansion and Gene Loss in Powdery Mildew fungi Reveal Tradeoffs in Extreme Parasitism. *Science* 330(6010): 1543-6. doi: 10.1126/science.1194573. Available online at www.sciencemag.org/content/330/6010/1543.long
30 The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475:189-95. doi:10.1038/nature10158. Available online at www.nature.com/nature/journal/v475/n7355/full/nature10158.html
31 Young N, et al (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480: 520-4. doi: 10.1038/nature10625. Available online at www.nature.com/nature/journal/v480/n7378/full/nature10625.html
32 Goodwin S, et al (2011). Finished Genome of the Fungal Wheat Pathogen *Mycosphaerella graminicola* Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. *PLoS Genet* 7(6): e1002070. doi:10.1371/journal.pgen.1002070. Available online at http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002070
33 Brassica rapa Genome Sequencing Project Consortium (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*. 43(10):1035-9. doi: 10.1038/ng.919.
34 Brenchley R et al (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491:705-10. doi: 10.1038/nature11650. Available online at www.nature.com/nature/journal/v491/n7426/full/nature11650.html
35 Groenen M, et al (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491:393-8. doi: 10.1038/nature11622. Available online at www.nature.com/nature/journal/v491/n7424/full/nature11622.html
36 Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635-41. doi: 10.1038/nature11119. Available online at www.nature.com/nature/journal/v485/n7400/full/nature11119.html
37 International Barley Genome Sequencing Consortium (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491:711-6. doi: 10.1038/nature11543. Available online at www.nature.com/nature/journal/v491/n7426/full/nature11543.html
38 Read B, et al (2013). Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* 499:209 13. doi: 10.1038/nature12221. Available online at www.ncbi.nlm.nih.gov/pubmed/23760476
39 Press release: 'Genome sequence for mother of ash dieback survival' TGAC.

40 TGAC releases new genetic data to combat ash dieback epidemic.

41 Peng X, Alföldi J, Gori K, Eisfeld AJ, Tyler SR, Tisoncik-Go J, Brawand D, Law GL, Skunca N, Hatta M, Gasper DJ, Kelly SM, Chang J, Thomas MJ, Johnson J, Berlin AM, Lara M, Russell P, Swofford R, Turner-Maier J, Young S, Hourlier T, Aken B, Searle S, Sun X, Yi Y, Suresh M, Tumpey TM, Siepel A, Wisely SM, Dessimoz C, Kawaoka Y, Birren BW, Lindblad-Toh K, Di Palma F, Engelhardt JF, Palermo RE, Katze MG. (2014). The draft genome sequence of the ferret (*Mustela putorius furo*) facilitates study of human respiratory disease. *Nat Biotechnol.* 32(12):1250-5. doi: 10.1038/nbt.3079. Available online at: www.nature.com/nbt/journal/v32/n12/full/nbt.3079.html

42 The International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome *Science* 345(6194). DOI: 10.1126/science.1251788. Available online at www.sciencemag.org/content/345/6194/1251788.full

43 Rivera D, Revale S, Molina R, Gualpa J, Puente M, Maroniche G, Paris G, Baker D, Clavijo B, McLay K, Spaepen S, Perticari A, Vazquez M, Wisniewski-Dyé F, Watkins C, Martínez-Abarca F, Vanderleyden J, Cassán F (2014). Complete genome sequence of the model rhizosphere strain *Azospirillum brasilense* Az39, successfully applied in agriculture. *Genome Announc*. 2(4):e00683-14. doi:10.1128/genomeA.00683-14. Available online at http://genomea.asm.org/content/2/4/e00683-14.full

44 Keane M, Craig T, Alföldi J, Berlin A, Johnson J, Seluanov A, Gorbunova V, Di Palma F, Lindblad-Toh K, Church G, de Magalhães J (2014). The Naked Mole Rat Genome Resource: facilitating analyses of cancer and longevity-related adaptations. *Bioinformatics*, 30(24), 3558–3560.

45 Press release: 'Yeast treasure-trove goes live'. BBSRC [Reference/webpage no longer available – July 2016]

46 Jose J. De Vega, Sarah Ayling, Matthew Hegarty, Dave Kudrna, Jose L. Goicoechea, Åshild Ergon, Odd A. Rognli, Charlotte Jones, Martin Swain, Rene Geurts, Chunting Lang, Klaus F. X. Mayer, Stephan Rössner, Steven Yates, Kathleen J. Webb, Iain S. Donnison, Giles E. D. Oldroyd, Rod A. Wing, Mario Caccamo, Wayne Powell, Michael T. Abberton, Leif Skøt. Red clover (*Trifolium pratense L.*) draft genome provides a platform for trait improvement. *Scientific Reports*, 2015; 5: 17394 DOI: 10.1038/srep17394. Available online: www.nature.com/articles/srep17394

47 'Sequencing power for every scale.' Illumina. documents/products/brochures/illumina-sequencing-portfolio-brochure-web-070-2011-003.pdf

48 Shiferaw, B., Smale, M., Braun, H-J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Sec.* 5:291–317. DOI 10.1007/s12571-013-0263-y

49 See, for instance, BBSRC Business Winter 2014: www.bbsrc.ac.uk/documents/bbsrc-business-winter-2014-wheat-pdf

50 Shirley, M.W., Ivens, A., Gruber, A., Madeira, A.M., Wan, K.L., Dear, P.H., Tomley, F.M. (2004) The Eimeria genome projects: A sequence of events. *Trends Parasitol.* 20:199–201.

51 About Genomic England and the 100,000 Genomes Project: www.genomicsengland.co.uk/the-100000-genomes-project/faqs/#