SERVICES & FACILITIES ANNUAL REPORT - FY April 2014 to March 2015

SERVICE	FUNDING	AGREEMENT	ESTABLISHED as S&F	TERM
NERC Biomolecular Analysis Facility, NBAF	PAYG & Block	F14/G6/48 (NBAF–B: R8/H10/61)	1998 (NBAF–S) 2005 (NBAF–E & NBAF–L) 2009 (NBAF–B & NBAF–W)	3 years

TYPE OF SERVICE PROVIDED:

NBAF provides advice on and access to a wide range of advanced molecular genetic, genomic, metabolomic and bioinformatic technologies to the UK environmental science community to address ecological or evolutionary questions. Since April 2009, service has been provided at four nodes: Birmingham (NBAF–B: metabolomics), Edinburgh (NBAF–E: sequencing and bioinformatics), Liverpool (NBAF–L: microarraying, sequencing and bioinformatics) and Sheffield (NBAF–S: genotyping and population genetics). Access to the Facility is organised centrally through competitive peer-reviewed proposals that are assessed by the independent members of the Steering Committee (application form at http://www.nbaf.nerc.ac.uk) to ensure that (i) only the best science is supported, (ii) access to more than one node is coordinated, and (iii) projects are followed through to dissemination of the results. Each node is embedded in a well-equipped and vigorous research environment that, together with continuing developments in equipment and training, ensures that the NBAF maintains a 'state-of-the-art' position. NBAF provides access to high-level capability, and the associated training, that are rarely available elsewhere. NBAF–S, and to a limited extent NBAF–B, are equipped to train and supervise researchers in undertaking their own analyses at the bench. NBAF–B supports metabolomic analyses using both mass spectrometry and NMR methods. At NBAF–S most studies require the development and implementation of highly polymorphic genetic markers, such as microsatellites and single nucleotide polymorphisms (SNPs). These and other polymorphisms (such as major histocompatibility loci) are then genotyped in large-scale studies using capillary and Illumina high-throughput sequencers or SNP-typing and qPCR platforms. NBAF–E offers experimental design consultancy, data generation and analysis based on high-throughput sequencing. NBAF–E uses ABI 3730 instruments for capillary sequencing and next-generation HiSeq 2500 and MiSeq instruments for whole-genome, transcriptome and reduced-representation sequencing, including targeted resequencing, amplicon sequencing and genotyping-by-sequencing, at any scale (from viruses to polyploid animals and plants). The service is pay-as-you-go (PAYG), and project-focussed bioinformatic training is also offered. NBAF–L offers sequencing, gene expression and bioinformatic services, particularly for environmental diversity through amplicon sequencing on long-read platforms, targeted resequencing of exons and reduced genomic regions, and gene expression analysis on microarray- and short-read (Illumina and Agilent) platforms. Building on dedicated bioinformatic expertise, NBAF–L provides an integrated experimental design and assay service, including statistical and network-based interpretation of results. NBAF–L is also PAYG.

ANNUAL TARGETS AND PROGRESS TOWARDS THEM

Capacity is defined by the availability of staff time, and all four nodes make >85% of funded staff time available to users, with the remainder allocated to R&D. Almost all projects this year have been accommodated according to the agreed schedule, with most slippage arising from delays in the arrival of users or their samples. One piece of equipment, the Illumina BeadXPress at NBAF–S, suffered from failures on the part of the manufacturer to provide working consumables. This led to significant delays in data delivery to a couple of users.

SCORES AT LAST REVIEW (each out of 5)			Date of Last Review:	2011
Need	**Uniqueness**	**Quality of Service**	**Quality of Science & Training**	**Average**
5	**4.5**	**5**	**5**	**4.88**

CAPACITY of HOST ENTITY FUNDED by S&F %	Staff (grade, fte): NBAF–S: DA Dawson (G7 100%), G Horsburgh (G7 100%), C Pagnier (G5 80%), H Hipperson (G7 100%), A Krupa (G7 20%); NBAF–B: U Sommer (G7 100%), J Sihra (G7 100%), J Engel (G7 100%); NBAF–E: K Gharbi (UOE8 20%), H Gunter (UOE7 40%), A Montazam (UOE5 100%), J Risse (UOE06 45%); NBAF–L: M Hughes (Res 8 50%) R Gregory (Res 7 50%), L Olohan (Res 7 100%), Y Fang (Res 7, 75%), C Hertz-Fowler (Res 9, 20%), L Parsons (Cler 5 30%)	Next Review (March)	Contract Ends (31 March)
		2016	2017

FINANCIAL DETAILS: CURRENT FY						
Total Resource Allocation £k	**Unit Cost £k**			**Capital Expend £k**	**Income £k**	**FCC £k**
	Unit 1	**Unit 2**	**Unit 3**			
1900.38		Variable		4.94		2027.77

FINANCIAL COMMITMENT (by year until end of current agreement) £k				
2011-12	**2012-13**	**2013-14**	**2014-2015**	**2015-2016**

STEERING COMMITTEE	Independent Members	Meetings per annum	Other S&F Overseen
NBAF SC	Chair (Prof R Nichols) + 7	1–2	None (one SC for 4 nodes).

APPLICATIONS: DISTRIBUTION OF GRADES (current FY — 2014/15) COMBINED													
	10	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**R***	**Pilot**
NERC Grant projects*	1	4	15	5	0	0	0	0	0	0	0	0	N/A
Other academic	0	4P fund	6 inc. 4P fund + 1P no fund	20 inc 17P no fund	8 inc. 6P no fund	0	0	0	0	0	0	0	32P
Students	0	1P fund	4P fund	11 inc. 5P no fund	4 inc. 3P no fund	2P no fund	1	0	0	0	0	0	15P
TOTAL 86	1	9	25	36	12	2	1	0	0	0	0	0	47P

PROJECTS COMPLETED (current FY – 2014/15) 157												
	10 (α5)	**9**	**8 (α4)**	**7**	**6 (α3)**	**5 (α2)**	**4**	**3 (α1)**	**2**	**1 (β)**	**0 (Reject)**	**Pilot**
NERC Grant projects*	4	20	47	10	1	0	0	0	0	0	0	N/A
Other Academic	1	6	18	5	0	0	0	0	0	0	0	20
Students	0	10	18	6	11	0	0	0	0	0	0	9

Project Funding Type (current FY – 2014/15) (select one category for each project)											
Grand Total	**Infrastructure**					**PAYG**					
	Supplement to NERC Grant *	**PhD Students**		**NERC Centre**	**Other**	**NERC Grant***	**PhD Students**		**NERC Centre**	**Other**
		NERC	**Other**				**NERC**	**Other**		
157	11	10.5	18	5	5.5	67	4.5	12	0	23.5
Project Funding Type (per annum average previous 3 financial years - 2011/2012, 2012/2013 & 2013/14)										
Grand Total	**Infrastructure**					**PAYG**				
	Supplement to NERC Grant *	**PhD Students**		**NERC Centre**	**Other**	**NERC Grant***	**PhD Student**		**NERC Centre**	**Other**
		NERC	**Other**				**NERC**	**Other**		
142	10.7	17.7	19.2	1	7	50.3	6	10.5	1.7	18

User type (current FY – 2014/15) (include each person named on application form)				
Academic	**NERC Centre**	**NERC Fellows**	**PhD Students**	**Commercial**
293	20	23	56	2
User type (per annum average previous 3 financial years - 2011/2012, 2012/2013 & 2013/14)				
Academic	**NERC Centre**	**NERC Fellows**	**PhD Students**	**Commercial**
266.7	15	20	61.7	1.3

OUTPUT & PERFORMANCE MEASURES (current year)										
Publications (by science area & type) (calendar year 2014)										
SBA	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**	**Grand Total**	**Refereed**	**Non-Ref/ Conf Proc**	**PhD Theses**
0	1	16	0	75	0	1	92	71	N/A	21

Distribution of Projects (by science areas) (FY 2014/15)							
Grand Total	**SBA**	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**
157	1	4	34	1	106	0	10

OUTPUT & PERFORMANCE MEASURES (per annum average previous 3 years)										
Publications (by science area & type) (Calendar years 2011, 2012 & 2013)										
SBA	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**	**Grand Total**	**Refereed**	**Non-Ref/ Conf Proc**	**PhD Theses**
0	0.3	22	0	90.3	0	1	113.6	72	18.3	23.3

Distribution of Projects (by science areas) (FY 2011/2012, 2012/2013 & 2013/14)							
Grand Total	**SBA**	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**
142	0.7	2	25.3	0	107.7	0	6.3

Distribution of Projects by NERC strategic priority (current FY 2014/15)							
Grand Total	**Climate System**	**Biodiversity**	**Earth System Science**	**Sustainable Use of Natural Resources**	**Natural Hazards**	**Environment, Pollution & Human Health**	**Technologies**
157	3.5	120.5	8	7	2.5	12	3.5

OVERVIEW & ACTIVITIES IN FINANCIAL YEAR

All NBAF nodes contributed to a workshop at the first International Environmental 'Omics Synthesis (iEOS) meeting in Liverpool (September 2014). Each node ran a hands-on data analysis training session, and these were well attended (>60 participants).

NBAF–B implemented a new LC-MS set-up for lipidomics for an NBAF project, adding to its existing direct infusion mass spectrometry lipidomics pipeline, which will increase its ability to annotate and identify lipids. The node also progressed its metabolite isolation and identification pipeline and purified a novel boron-containing marine bacterial compound. NBAF–B received funding for an ongoing project to automate the preparation of metabolomics samples using tissue homogenisation and liquid handling robotics.

NBAF-E is part of Edinburgh Genomics, which runs a rich series of training courses for NERC and other interest group researchers. In 2014-15, Edinburgh Genomics offered courses in RNA-Seq data analysis (offered 8 times; fully booked for each course), Introduction to LINUX for biologists (offered 3 times, fully booked), Introduction and Advanced Partek Analysis Suite (two sessions), and Introduction to Python for Biologists (week-long course, fully booked), in addition to the RAD-Seq Analysis workshop offered to iEOS. Edinburgh Genomics courses are promoted to the NERC community through the NBAF mailing list.

NBAF–L ran a gene expression workshop jointly with The Physiological Society at Liverpool in March 2015. It ran an introduction to single-molecule sequencing on the Pacific Biosciences SMRT sequencer (April 2014) and to the Nanostring nCounter system (April 2014).

NBAF–S ran several training courses during the year, with many users staying on immediately afterwards in Sheffield to complete their projects in the laboratory. The node worked with the University of Sheffield and the NHS to establish a local core sequencing facility and this has provided ready access to Illumina MiSeq equipment, in particular. The MiSeq is taking over some of the work of the ABI machines. NBAF–S also won NERC capital funding to upgrade its SNP typing and qPCR facilities, to be purchased in the coming year.

Significant staff changes at NBAF–W led to a review of the node's activities and the Steering Committee, in consultation with CEH, took the decision to close this node and to provide bioinformatic support via the other nodes. No additional work was allocated to NBAF–W and all projects in hand were completed by March 2015 or else alternative support was arranged.

SCIENCE HIGHLIGHTS

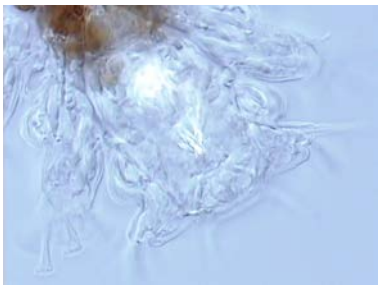
^{1}H NMR Metabolomics reveals contrasting response by male and female mussels exposed to reduced seawater pH, increased temperature, and a pathogen (Ellis *et al.* 2014 *Environ. Sci. Technol.* 48, 7044–52; 10.1021/es501601w [Journal Impact Factor: 5.48])

Ocean acidification (OA) is occurring against a background of warming and an increasing occurrence of disease outbreaks, posing a significant threat to marine organisms, communities, and ecosystems. ^{1}H NMR spectroscopy was used to investigate the response of the blue mussel, *Mytilus edulis*, to reduced seawater pH and increased temperature, followed by a subsequent pathogenic challenge. Analysis of the metabolome revealed significant differences between male and female organisms and different responses to environmental stress. This study has important implications for the interpretation of metabolomic data in mussels, as well as for understanding the impact of environmental stress in marine invertebrates in general. NBAF–B contributed to planning, instrumental analysis and data interpretation.

The evolution of sex ratio distorter suppression affects a 25-cM genomic region in the butterfly *Hypolimnas bolina* (Hornet *et al.* 2014 *PLoS Genetics* 10, e1004822; 10.1371/journal.pgen.1004822 [JIF: 8.50])

The sex ratio of the offspring produced by an individual can be an evolutionary battleground. In many arthropod species, maternally inherited microbes selectively kill male hosts, and the host may in turn evolve strategies to restore the production or survival of males. When males are rare, the intensity of selection on the host may be extreme. Greg Hurst and colleagues recently observed one such episode, in which the population sex ratio of the butterfly *Hypolimnas bolina* shifted from 100 females per male to near parity, through the evolution of a suppressor gene. By using genome sequencing, they were able to show that genetic diversity was massively reduced across a broad region of the genome as the suppressor spread rapidly through the population. Their data show how genomes can be shaped by strong episodes of selection in the natural environment. Work was conducted jointly at NBAF–E and –L.

Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes (Fonseca *et al*., 2014 *Global Ecology and Biogeography* 23, 1293–1302; 10.1111/geb.12223 [JIF: 7.2])



Traditional methods for characterization of species diversity within environmental samples are laborious and limit our ability to perform ecological surveys. Fonseca *et al.* demonstrated the potential of new techniques in the characterization of DNA within environmental samples (eDNA) that allows much faster and deeper quantification of diversity within the coastal environment. These are enabled by next-generation sequencing methods. The authors highlight the potential of this powerful technique to uncover ecological diversity across a range of ecological scales. These techniques are now starting to be widely employed by both academic researchers and environmental consultancies. This work was supported by NBAF–L.

Using social parasitism to test reproductive skew models in a primitively eusocial wasp (Green, Cant & Field 2014 *Proceedings of the Royal Society B* 281, 1206**;** 10.1098/rspb.2014.1206; [JIF 5.68])

Remarkable variation exists in the distribution of reproduction (skew) among members of cooperatively breeding groups, both within and between species. Reproductive skew theory has provided an important framework for understanding this variation. Two models have been routinely tested in the primitively eusocial Hymenoptera – concessions models, which assume complete control of reproduction by a dominant individual, and tug-of-war models, which assume ongoing competition among group members over reproduction – but previous work provided little support for either model. The authors used a direct test of concessions and tug-of-war models in the paper wasp *Polistes dominulus* by exploiting pronounced changes in relatedness and power structures that occur following replacement of the dominant by a congeneric social parasite. Comparisons of skew in parasitized and unparasitized colonies were consistent with a tug-of-war over reproduction within *P. dominulus* groups, but provided no evidence for reproductive concessions. NBAF–S supported microsatellite marker development and genotyping. (Image of *Polistes* by Jasper van Huesden.)

Rapid convergent evolution in wild crickets (Pascoal *et al.* 2014 *Current Biology* 24, 1367–1374; 10.1016/j.cub.2014.04.053 [JIF: 9.9])

The earliest stages of convergent evolution are difficult to observe in the wild, limiting our understanding of the incipient genomic architecture underlying convergent phenotypes. To address this, we capitalized on a novel trait, flatwing, that arose and proliferated at the start of the 21st century in a population of field crickets (*Teleogryllus oceanicus*) on the Hawaiian island of Kauai. Flatwing erases sound-producing structures on male forewings. Mutant males cannot sing to attract females, but they are protected from fatal attack by an acoustically orienting parasitoid fly (*Ormia ochracea*). Two years later, the silent morph appeared on the neighbouring island of Oahu. The authors tested two hypotheses for the evolutionary origin of flatwings in Hawaii: (1) that the silent morph originated on Kauai and subsequently introgressed into Oahu and (2) that flatwing originated indepenently on each island. Standard genetic crosses confirmed that the mutation segregates as a single-locus, sex-linked Mendelian trait on both islands. However,

genome-wide scans using RAD-seq recovered almost completely distinct markers linked with flatwing on each island. The patterns of allelic association with flatwing on either island reveal different genomic architectures consistent with the timing of two separate mutational events on the X chromosome. Divergent wing morphologies linked to different loci have thus caused an identical behavioural outcome – silence – illustrating the power of selection to rapidly shape convergent adaptations from distinct genomic starting points. NBAF–E carried out library preparation and Illumina sequencing, and provided bioinformatics support.

Cooperative investment in public goods is kin-directed in communal nests of social birds (Van Dijk *et al*. 2014 *Ecology Letters*, 17, 1141–1148; 10.1111/ele.12320 [JIF: 17.9])

The tragedy of the commons predicts social collapse when public goods are jointly exploited by individuals attempting to maximize their fitness at the expense of other social group members. However, animal societies have evolved many times despite this vulnerability to exploitation by selfish individuals. Kin selection offers a solution to this social dilemma, but in large social groups mean relatedness is often low. Sociable weavers (*Philetairus socius*) live in large colonies that share the benefits of a massive communal nest, which requires individual investment for construction and maintenance. Here, we show that despite low mean kinship within colonies, relatives are spatially and socially clustered and that nest-building males have higher local relatedness to other colony members than do non-building males. Alternative hypotheses received little support, so we conclude that the benefits of the public good are shared with kin and that cooperative investment is, despite the large size and low relatedness of these communities, kin directed. NBAF–S supported microsatellite development and genotyping. (Image of communal nest provided by Rene van Dijk.)

FUTURE DEVELOPMENTS/STRATEGIC FORWARD LOOK

Metabolomics

After ten years of operation, NBAF–B's Thermo LTQ FT Ultra mass spectrometer has been retired because of high operating costs and an increasing number of hardware failures. Through BBSRC (50%) and University of Birmingham (50%) funding, a Bruker 7T FT-ICR Solarix mass spectrometer is now being installed. One future (and much-needed) development will be writing new software (by Dr Engel) to integrate both this new Bruker instrument and the relatively recently purchased Thermo Q Exactive mass spectrometer into NBAF's data processing and analysis workflows, to allow both direct infusion MS and LC-MS. In addition, we will re-optimise the direct infusion "SIM stitching" analytical method on both of these mass spectrometers (by Dr Sommer). A further development planned for 2015–16 is to implement nanoflow LC-MS methods on the Thermo Q Exactive to provide customers with a more sensitive alternative to our existing UHPLC-MS for small sample sizes and less abundant compounds, such as signalling molecules.

Ultrahigh-Throughput Sequencing

2014–15 saw a continued increase in the number and diversity of sequencing projects, conducted at NBAF–E and NBAF–L, funded through PAYG. These projects include *de novo* sequencing of transcriptomes from a wide range of taxa, RAD-Seq reduced representation genome-wide genotyping, genome resequencing, sequence capture, metagenomics, metabarcoding and digital gene expression. Within Edinburgh Genomics, research and development activities have focussed on proving production protocols for double-digest RAD sequencing library production, and also developing custom targeted sequence capture protocols fit for purpose for the ecological and environmental genomics communities. We are seeing an increase in the number of *de novo* sequencing projects of organisms of environmental interest, and that the sizes of genomes tackled are becoming larger as users now exploit the technology to produce assembled genomes of target plants and animals. We have also recently expanded our range of applications to include epigenetics (bisulphite sequencing and ChIP-Seq).

Both NBAF–E and NBAF–L are embedded within larger sequencing facilities – The Centre for Genomic Research (CGR) at Liverpool and Edinburgh Genomics at Edinburgh – which receive substantial funding from other research councils, charities and industry. This provides NERC-funded research with substantial advantages: economies of scale in pricing, access to leading-edge technologies across all major sequencing platforms, and a breadth of core technical and bioinformatic expertise in next-gen sequencing. The CGR at Liverpool is making a substantial investment in single-cell genomics and in synthetic biology, which will aid researchers particularly in understanding microbial communities in the natural environment and in exploiting novel genes for bioindustry.

NBAF is committed to continuing to enable access to the latest sequencing technologies. A number of technologies on the horizon present opportunities for NERC users to perform innovative sequencing projects. (I) Longer sequence reads. NBAF–L has acquired a PacBio SMRT sequencer, funded by BBSRC, NERC and Unilever, which enables reads of >10 kb. Edinburgh Genomics has been at the forefront of testing Oxford Nanopore minION long read technology, including developing new software tools for data analysis, and exploring its utility in rare event/pathogen detection, and in scaffolding genome assemblies to improve genome contiguity. Both technologies aid the assembly of novel genomes and the characterisation of epigenetic modifications, and nanopore sequencing has the potential to allow portable sequencing in the field. (II) More samples. There is increasing demand from NERC users to analyse more individuals, particularly from the very successful RAD-tagging platform at NBAF–E, but also from RNAseq, sequence capture and whole-genome bacterial sequencing. To accommodate this demand, NBAF–E and NBAF–L have invested heavily in robotic library preparation and in shared, compatible laboratory information management software (LIMS) to track all aspects of projects. (III) Lower sample volume. New questions can be answered by finer-scale resolution of samples within environmental mixtures. Both NBAF–E and –L continue to trial kits that allow less input material and NBAF–L is setting up a dedicated single-cell genomics laboratory, to include cell-sorting, laser capture and automated sample preparation. (IV) Greater volume of sequence. While reduced-representation approaches to genotyping-by-sequencing are still very valid, and increasingly accessed by the NERC community (see below), as per-base costs drop, NERC science is moving to full genome sequencing of target organisms for population genomic analyses, which will allow researchers to identify loci under selection in the natural environment and to infer fine-scale demographic processes. Both NBAF–E and –L continue to invest heavily in sequencing machines to accommodate this demand, and coordinate closely with each other and with other UK academic sequencing facilities to share loads during machine downtime or spikes in demand. (V) More computation. With the growth in data generation capacity, the need has grown for high-performance computing to cope with assembly, mapping and archiving of increasing volumes of data. The host facilites of NBAF–E and NBAF–L continue to invest heavily in high-quality, high-resilience and high-specification compute systems and in bioinformatic staff that can support the full range of bioinformatic analyses resulting from sequencing experiments.

Gene expression profiling. To understand the response of organisms to their environment, NBAF offers fabricated microarray and RNAseq solutions to gene expression, with advice on the most appropriate platform and design based on questions addressed and cost effectiveness. Increasingly, gene expression experiments are coupled with *de novo* genome sequencing to provide a robust reference for analyses. Bioinformatic and statistical support are provided either through PAYG of NBAF staff time, or by training of users. NBAF–L also offers NanoString detection of gene expression of targets using PCR-free detection.

Species diversity and eDNA. A growing area of NERC science, comprehensively supported by NBAF–E, NBAF–L and NBAF–S, is using sequencing methods, including eDNA, to characterise species diversity from the natural environment. Amplicon sequencing of genetic barcodes, such as 16S, is offered on Illumina platforms and we now are able to offer longer reads on the PacBio platforms. Considerable development has been deployed to provide validated analysis pipelines to characterise taxa and quantify sample diversity. Future work will improve analytical ability for shotgun sequencing of complex communities to characterise gene function. eDNA analyses include using microsatellite markers to identify individuals and metabarcoding to identify the composition of diets from faecal samples.

Population genomics. Users are using increasing numbers of genetic markers, derived from genomic data, in population genetic studies. Microsatellite markers, now obtained by high-throughput sequencing, are still in demand and the marker of choice for many questions, especially including parentage analyses in behavioural ecological studies. Several studies in the last year exploited SNP data using the Illumina BeadXpress system at NBAF–S, though this platform is becoming obscolescent and a replacement (the LGC SNPLine) will shortly be purchased with NERC capital grant and University of Sheffield support to NBAF–S's host laboratory. Multiple population genomic studies used RNAseq, dRAD or ddRAD analyses to identify SNPs.

There is increasing collaboration among NBAF–E, –L and –S in delivering support to users. For example, projects that used SNP data from reduced representational sequencing, or that required metabarcoding, were delivered by both NBAF–E and –S. Users that require close support and training at the bench, from DNA extraction onwards, are supported in Sheffield, and the sequencing is then conducted in Sheffield (MiSeq) or Edinburgh or Liverpool (HiSeq). Sheffield, in turn, provides thorough support through its specialist provision for population genomics data analysis. More sophisticated users, who have appropriate molecular skills and so only require sequencing or more limited bioinformatic support, work directly with Edinburgh or Liverpool.