SERVICES & FACILITIES ANNUAL REPORT - FY April 2019 to March 2020

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)	N/A (Ends 31 March 2021)

TYPE OF SERVICE PROVIDED:

NBAF provides fully-supported access to advanced molecular genetic, genomic, metabolomic and bioinformatic technologies to the UK environmental science community to address ecological or evolutionary questions. Support is provided at three nodes: Edinburgh (NBAF–E: sequencing and bioinformatics), Liverpool (NBAF–L: sequencing and bioinformatics) and Sheffield (NBAF–S: population genomics and environmental DNA [eDNA]). Access is organised centrally through competitive peer-reviewed proposals, assessed by the independent members of the Steering Committee 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, makes 'state-of-the-art' services available to the NERC community. NBAF provides access to high-level capability, and the associated training, that are rarely available elsewhere. NBAF–E supports whole-genome, transcriptome and reduced-representation sequencing, including targeted resequencing, amplicon sequencing and genotyping-by-sequencing, at any scale, using short-read (Illumina NovaSeq and MiSeq) and long-read (PacBio Sequel, ONT Promethion) instruments. NBAF–L offers sequencing, gene expression and bioinformatic services, particularly for environmental diversity through amplicon sequencing on long-read (PacBio) platforms, targeted resequencing of exons and reduced genomic regions, and gene expression analysis on short-read (Illumina) platforms. NBAF–S is equipped to train and supervise researchers (mostly PhD students) in their own analyses at the bench. At NBAF–S, most studies require metabarcoding (for biodiversity, dietary or microbiome analyses) or genotyping (microsatellites, ddRAD, single nucleotide polymorphisms [SNPs], target amplicons or telomeres) methods. Samples are prepared and analysed using Illumina, PacBio or ABI sequencers, SNP typing or qPCR platforms. All NBAF nodes provide integrated experimental design, data generation and bioinformatic analysis; NBAF–L also offers statistical and network-based interpretation of results. NBAF provides a wide diversity of training courses in genomics and population genetics, as well as project-specific wet lab and bioinformatics training. Access to NBAF–E and NBAF–L is "pay-as-you-go" (PAYG), to support grant-funded projects, but NBAF coordinates an annual, fully-funded Pilot Project scheme to introduce NERC researchers to next-generation sequencing approaches in ecological and environmental science.

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 any remainder allocated to R&D. Projects have run largely as scheduled, except where users have failed to submit adequate samples on time.

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%), K Maher (G7 100%), C Pagnier (G5 80%), R Tucker (G6 20%); NBAF–E: U Trivedi (UOE7 40%), M Arno (UOE07 15%), R Talbot (UoE7 30%), Alison Morris (UoE5, 100% to 13/10/2019), Nathan Medd (UoE7, 70% from 1/10/19); NBAF–L: K Jackson (Res 9 20%), L Parsons (Cler 5 30%), R Gregory (Res 7 20%), J Kenny (Res 8 10%), C Owen (Res 6 100%), X Lui (Res 7 80%).	Next Review (March)	Contract Ends (31/3)
N/A		–	2021

FINANCIAL DETAILS: CURRENT FY						
Total Resource Allocation	Unit Cost £k			Capital Expend	Income £k	FCC
	Unit 1	Unit 2 Variable	Unit 3	4.94		£2,027.8
FINANCIAL COMMITMENT (by year until end of current agreement) £k						
2018-19	**2019-20**	**2020-21**				

STEERING COMMITTEE	Independent Members	Meetings per annum	Other S&F Overseen
NBAFSC	Chair +6	1-2	None (one SC for 3 nodes)

APPLICATIONS: DISTRIBUTION OF GRADES (current FY — 2019/20)													
	10	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**R***	**Pilot**
NERC Grant projects*	2	6	3	0	0	0	0	0	0	0	0	0	N/A
Other academic	0	0	0	0	0	0	1	0	0	0	0	0	31
Students	0	0	5	5	2	3	2	0	0	0	0	0	15
TOTAL	2	6	8	5	2	3	3	0	0	0	0	0	46
APPLICATIONS: DISTRIBUTION OF GRADES (per annum average previous 3 financial years — 2016/2017 & 2017/2018 & 2018/2019)													
	10	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**R***	**Pilot**
NERC Grant projects*	1.0	5.3	6.3	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	N/A
Other academic	0.0	0.0	1.0	1.3	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.3	21.3
Students	0.0	0.7	6.3	5.3	3.3	1.0	0.3	0.0	0.0	0.0	0.0	1.7	15.3
TOTAL	1.0	6.0	13.7	7.3	4.7	1.0	0.3	0.0	0.0	0.0	0.0	2.0	36.7

PROJECTS COMPLETED (current FY – 2019/20) 110												
	10	**9**	**8 (α4)**	**7**	**6 (α3)**	**5 (α2)**	**4**	**3 (α1)**	**2**	**1 (β)**	**0**	**Pilot**
NERC Grant projects*	6	17	27	2	0	0	0	0	0	0	0	
Other Academic	0	0	1	3	0	0	0	0	0	0	0	18
Students	0	0	11	11	7	0	0	0	0	0	0	7

Project Funding Type (current FY – 2019/20) (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**		
110	**1**	**13**	**16**	**0**	**4**	**43**	**8**	**7**	**0**	**18**
Project Funding Type (per annum average previous 3 financial years — 2016/2017 & 2017/2018 & 2018/2019)										
Grand Total	**Infrastructure**					**PAYG**				
	Supplement to NERC Grant *	**PhD Students**		**NERC Centre**	**Other**	**NERC Grant***	**PhD Student**		**NERC Centre**	**Other**
		NERC	**Other**				**NERC**	**Other**		
104.2	3.0	14.2	18.8	1.0	8.0	43.5	0.5	4.5	0.0	10.7

User type (current FY – 2019/20) (include each person named on application form)				
Academic	**NERC Centre**	**NERC Fellows**	**PhD Students**	**Commercial**
267	1	10	41	3
User type (per annum average previous 3 financial years — 2016/2017 & 2017/2018 & 2018/2019)				
Academic	**NERC Centre**	**NERC Fellows**	**PhD Students**	**Commercial**
206	18	8	30	2

OUTPUT & PERFORMANCE MEASURES (current year)										
Publications (by science area & type) (calendar year 2019)										
SBA	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**	**Grand Total**	**Refereed**	**Non-Ref/ Conf Proc**	**PhD Theses**
0	0	9.3	0	72.3	0	1.3	83	55	14	14

Distribution of Projects (by science areas) (FY2019/20Y)							
Grand Total	**SBA**	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**
110	0.5	6.25	12.5	0.75	86.8	0	3.2

OUTPUT & PERFORMANCE MEASURES (per annum average previous 3 years)										
Publications (by science area & type) (Calendar years 2016, 2017 & 2018)										
SBA	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**	**Grand Total**	**Refereed**	**Non-Ref/ Conf Proc**	**PhD Theses**
0	1.1	17.6	0	87.5	0	3.8	110	72	18	19.7

Distribution of Projects (by science areas) (2016/2017 & 2017/18 & 2018/2019)							
Grand Total	**SBA**	**ES**	**MS**	**AS**	**TFS**	**EO**	**Polar**
104	0	4.2	17.8	0	79.3	0	2.7

Distribution of Projects by NERC strategic priority (current FY 2019/20)							
Grand Total	**Climate System**	**Biodiversity**	**Earth System Science**	**Sustainable Use of Natural Resources**	**Natural Hazards**	**Environment, Pollution & Human Health**	**Technologies**
110	15.7	80	0	5.6	5.9	2.6	0

OVERVIEW & ACTIVITIES IN FINANCIAL YEAR (2019/20):

NBAF-E is delivered from within Edinburgh Genomics (EdGe), one of the largest genomics facilities in the UK. It has been a year of change for EdGe. At the end of January 2020, EdGe consolidated all its operations in a single laboratory, closing down the high-throughput whole-genome sequencing (WGS) facility based at the Roslin Institute. The restructured EdGe, operating from its facility at the Ashworth laboratories, was launched in February 2020. This is an agile facility with an experienced team that covers a broad range of short- and long-read sequencing technologies, genomics applications, bioinformatics data analyses and training courses. The support offered by the facility extends from experimental design through to assistance with publication, and includes long experience with "difficult" – especially low -quantity – sample sets. In April 2019, NBAF-E launched its Oxford Nanopore Technologies PromethION early access programme, which has delivered 19 early-access projects from a wide range of species, as well as microbiomes. The PromethION service is now available to any user. The facility has also run a number of major research and development projects, including optimisation of reduced representation methylation sequencing, Reverse Metagenomics (RevMeta) and Hi-C. The node had to temporarily close to users during the COVID-19 pandemic in March 2020.

NBAF-L is delivered from within the Centre for Genomics Research (CGR). In 2019 it acquired an Illumina Novaseq platform (through NIHR funding), which allows it to offer high-throughput and flexible short-read sequencing. The Novaseq is now fully embedded in our laboratory workflows and has been deployed to support NERC users in RNAseq, genome resequencing, metagenomic and single-cell genomic work. We have also embedded new epigenetic workflows and successfully miniaturised gDNA library preparation to reduce costs. We continued to support long-read sequencing through our Pacific Biosciences Sequel platform, which supports NERC scientists in providing de novo transcriptome references and isoform analysis through the IsoSeq protocol, and in whole-genome assembly. Considerable effort in 2019/20 has been made to develop bioinformatic workflows for several applications, including metagenetics/metagenomics, gene expression and whole-genome assembly and annotation. This allows us to provide a robust service and liberates staff time for more bespoke analyses. NBAF-L ran short courses in gene analysis and metagenomic analysis of communities, which were well received and over-subscribed. At the end of March 2020, normal NBAF work stopped and all CGR staff and equipment (including NBAF-funded staff, with NERC approval) were deployed for genomic surveillance using rapid sequencing of SARS-CoV-2 samples as part of the COVID-19 UK Genome Consortium.

NBAF-S is hosted by the Sheffield Molecular Ecology Laboratory. In 2019–20 the node supported microsatellite typing, ddRAD analyses, metabarcoding, MHC analyses and bird sexing. Microsatellite projects involved studies of population structure/ kinship/ landscape genetics of pine marten, snow leopard, red mason bee, long-tailed tit, monk parakeet, rhinoceros and camel. New projects requiring metabarcoding techniques included the analysis of diet from faecal samples in hedgehogs and badgers, pygmy shrew and greater white-toothed shrew, and from stomach contents in hoverflies, biodiversity analyses of arthropods in semi-natural and plantation oak woodlands, and microbial community analyses within pitcher plants. Metabarcoding techniques were also applied to MHC analyses in Seychelles warblers. ddRAD-based population genomic studies included studies on hybridisation and genetic structure in Morelet's crocodiles, assessing gene flow and species boundaries in sympatric *Myotis* bats in Mexico, and assessing genetic diversity and local adaptation in the mountain ringlet butterfly. NBAF-S ran two courses this year on metabarcoding. The NBAF-S lab closed on the 16 March 2020 due to the Covid-19 pandemic. Staff adapted to provide remote training and assist analyses and manuscript preparation. Technical staff were redeployed to sequence SARS-CoV-2 samples as part of the COVID-19 UK Genome Consortium.

NBAF funded 18 Pilot Grant projects in October 2019, enabling new users (including PhD projects) to gain access to NBAF-E and NBAF-L, which are otherwise only accessible via PAYG.

SCIENCE HIGHLIGHTS

In 2019–20 NBAF serviced 110 projects, including 44 supported by NERC grants; ran courses; supported 267 academics and 41 students; and in 2019 supported the work reported in 55 refereed publications.

Several examples from these publications are described below.

Meiotic drive reduces egg-to-adult viability in stalk-eyed flies (Finnegan *et al.* 2019 *Proceedings of the Royal Society London B* **286**, 20191414). NBAF supported microsatellite genotyping and data analysis.

Several species exhibit the phenomenon of Sex-Ratio (SR) meiotic drive, a selfish genetic element located on the X-chromosome that causes dysfunction of Y-bearing sperm, and one that has the potential to be exploited in the control of harmful insect vectors such as some mosquitoes. SR is transmitted to up to 100% of offspring, causing extreme sex-ratio bias. SR in several species is found in a stable polymorphism, suggesting that there must be strong frequency-dependent selection resisting its spread. This study investigated the effect of SR on female and male egg-to-adult viability in the Malaysian stalk-eyed fly, *Teleopsis dalmanni*. SR meiotic drive in this species is old, and appears to be broadly stable at a moderate (*ca* 20%) frequency. Large-scale controlled crosses were used to estimate the strength of selection acting against SR in female and male carriers. SR was found to reduce the egg-to-adult viability of both sexes. In females, homozygous females experienced greater reduction in viability and the deleterious effects of SR were additive. The male deficit in viability was not different from that in homozygous females. The study contributed to our understanding of how these reductions in egg-to-adult survival maintain the SR polymorphism in this species. (Photo credit: Kiran Lee.)

Individual variation in early-life telomere length and survival in a wild mammal (van Lieshout *et al.* 2019, *Molecular Ecology* **28**, 4152–4165). NBAF supported telomere length analysis using quantitative PCR methods.

Individual variation in survival probability due to differential responses to early-life environmental conditions is important in the evolution of life histories and senescence. A biomarker allowing quantification of such individual variation, and that links early-life environmental conditions with survival by providing a measure of the conditions that have been experienced, is telomere length. This study examined telomere dynamics in 24 cohorts of European badgers (*Meles meles*). It found a complex cross-sectional relationship between telomere length and age, with no apparent loss over the first 29 months, but with both decreases and increases in telomere length at older ages. Overall, there was low within-individual consistency in telomere length over individual lifetimes. Unexpectedly, some individuals showed reliable increases in telomere length. There was evidence that early-life telomere length predicts lifespan due to an effect on survival to adulthood (≥1 year old), but not on adult survival probability. These results showed that the relationship between early-life telomere length and lifespan was driven by conditions in early-life, where early-life telomere length varied strongly among cohorts. Our data provide evidence for associations between early-life telomere length and individual life history, and highlight the dynamics of telomere length across individual lifetimes due to individuals experiencing different early-life environments. (Photo credit: Inigo Montes.)

Parallel evolution of complex centipede venoms revealed by comparative proteotranscriptomic analyses (Jenner *et al*. 2019, *Molecular Biology and Evolution* **36**, 2748–2763). NBAF supported transcriptome sequencing.

The evolution of venom is an ancient adaptation to the predatory lifestyle of centipedes. Until now, venom evolution has been studied in only one of the five orders of centipede. Using transcriptome sequencing for non-model species, Jenner *et al*. were able to uncover a striking diversity of venom components comprising 93 phylogenetically distinct protein and peptide families. They were able to uncover the processes of gene duplication leading to this diversity and found that the majority, 67, of these families were found in single orders. Integration of transcriptomes with proteomic analysis was central to their approach in assigning function to gene families. (Photo credit: Eric Guinther,

Wikimedia)

Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. (George *et al.* 2019 *Nature Communications* **10**, 1107) NBAF supported metabarcoding.

Biodiversity within soil accounts for around 25% of global biodiversity, yet the large-scale distribution patterns of this biodiversity are very poorly understood. Here, George *et al*. used rapid, molecular approaches to quantity this diversity across UK landscapes in a major survey of 436 sites. They showed that animal microbial diversity followed markedly different patterns; animal biodiversity was driven by intensive land use, whereas microbial biodiversity was driven by soil properties. This benchmark survey of UK soil, using genomic methods, provides a solid framework to understand and value this important component of UK biodiversity. (Photo credit: Colin Kinear, Wikimedia.)

Gradients in richness and turnover of a forest passerine's diet prior to breeding: A mixed model approach applied to faecal metabarcoding data (Shutt et al. 2020 *Molecular Ecology*, **29**, 1199–1213) NBAF supported metabarcoding and bioinformatic analysis.

Generalist insectivorous bird species may feed on a range of different prey items, though quantifying variation in prey type and feeding preferences remains a substantial challenge. In this study, Shutt et al. used faecal DNA metabarcoding to assess how diet of blue tits varies across a transect of 39 sites in Scotland. They showed that most prey items are rare, and that diet composition shows substantial change through the feeding season. This study provides useful insights into temperate foodwebs, and shows how DNA metabarcoding can provide a robust framework for estimating how diet composition changes in both space and time. (Photo credit: Francis Franklin, Wikimedia.)

Lack of long-term acclimation in Antarctic encrusting species suggests vulnerability to warming (Clark et al. 2019 *Nature Communications*, **10**, 3383) NBAF supported RNA-seq and metabarcoding.

Polar marine invertebrates are predicted to experience major population declines due to warming oceans. Understanding how individual species' respond to temperature changes in an experimental setting may allow us to predict wider responses in ocean systems. Clark *et al*. investigated the responses of spatially dominant calcified marine worms in artificial heating experiments using RNA-Seq and expression profiling. They contrasted these results with co-occurring prokaryotic bacteria communities investigated with amplicon sequencing. Marine worms showed significant stress responses in warmed environments, with this response contrasting to bacteria that showed no difference in community composition with temperature. These results show different ecosystem changes in response to a changing climate. (Photo credit: Lovell and Libby Langstroth.)

FUTURE DEVELOPMENTS/STRATEGIC FORWARD LOOK

During 2019/20, recommissioning for a new environmental omics facility took place under an open competition. NBAF will therefore be ramped down in 2020–2022 as it completes its existing projects. Its successor, the NERC Environmental Omics Facility (NEOF), will begin operation in October 2020. It will be provided by the Universities of Liverpool and Sheffield, and will be managed by BGS.

The recommissioning provided an opportunity to reimagine the omics support required to support world-class NERC science. A central principle is to add value to the work of NERC scientists, and our ambition is for NERC science to be world-leading in environmental omics. This involves not just providing a service, but helping to build a community of skilled scientists with the capability to apply cutting-edge technologies to their own research. Relative to NBAF, the new NEOF facility will:

- Greatly expand training through research stays, a dedicated training suite, bespoke courses, on-line provision and one-to-one mentorship to build, develop and embed omics skills within the community.
- Increase the resources available to pilot and student projects to build capacity in the application of new techniques and skills within the community.
- Provide robust experimental design and intensive, bespoke and continuous support to all projects, from initiation to publication, and build capacity into the community through the delivery of workshops and showcase events.
- Expand capability beyond genomics into quantitative proteomics, metabolomics and phenomics, to integrate multiple types of omics and phenotypic data.
- Extend omic analysis to single cells using single-cell RNAseq, high-sensitivity MS-based proteomics and spatially resolved metabolomics to reveal cellular heterogeneity.
- Provide bioinformatic support to all projects with provision expanded to include the integration of meta'omic and ecological data, and machine-learning.
- Provide data stewardship and metadata integration for all projects under FAIR principles.

The University of Liverpool will provide expertise and technology in the areas of genomics, proteomics and metabolomics. The University of Sheffield will provide a research hotel for laboratory training, sample preparation and environmental DNA analysis, and lead on phenomics to supplement other omic approaches. The user perspective will be to have a single point of entry to the facility. Projects will be able to receive support from both nodes; for example for PhD students to prepare samples at Sheffield, which are then sequenced at Liverpool. Bioinformatics training will be provided across the facility, building on expertise in the research domains found at each node. We wish to grow the research community in new areas, notably proteomics, metabolomics and phenomics. The key mechanism to achieve this will be through Highlight Topics running for 12–18 months; these will build an understanding of the potential of the technology, support pilot studies to test ideas, and train researchers.

NEOF will partner with the Darwin Tree of Life Programme (ToL), based at the Sanger Institute, which is producing high-quality reference genomes for species from the British Isles, and NEOF is discussing with ToL how to avoid duplication of effort in generating genomes. The aim is to maximise the value of the ToL genomes using NERC-funded studies to explore ecological and environmental questions using a variety of omic approaches within NEOF. We are discussing with ELIXIR how to curate environmental omic data under FAIR principles by making it as easy as possible for users to gather metadata at the outset of a study.