

SERVICES & FACILITIES ANNUAL REPORT - FY April 2010 to March 2011

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

TYPE OF SERVICE PROVIDED:

NBAF provides 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 five nodes: Birmingham (NBAF-B: metabolomics), Edinburgh (NBAF-E: sequencing and bioinformatics), Liverpool (NBAF-L: microarraying, sequencing and bioinformatics), Wallingford (NBAF-W: 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.nbafe.nerc.ac.uk>) to ensure that (i) only the best science is supported, (ii) that access to more than one node is coordinated, and (iii) that 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. The 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 of highly polymorphic genetic markers, usually microsatellites or single nucleotide polymorphisms. These and other markers (such as amplified fragment length polymorphisms and Major Histocompatibility Locus reference strand mediated conformation analysis) are then genotyped in large-scale studies using ABI automated capillary sequencers. NBAF-E offers capillary sequencing of genomic and cDNA fragments and clones using ABI 3730 instruments and next-generation high-throughput sequencing of genomes and transcriptomes, and generation of population genomic marker data, using Illumina GAIIX and HiSeq, and Roche 454 titanium instruments; dedicated large-scale bioinformatic analysis is also provided. The service is pay-as-you go, and bioinformatic training is also offered. NBAF-L offers sequencing and gene expression services. Genome, exome and transcriptome sequencing ARE offered on its 454 platform, with associated bioinformatic processing. Gene expression analysis is offered on microarray and short-read SOLiD platforms. NBAF-L provides an integrated experimental design and assay service, including statistical- and network-based interpretation of results. Users either submit samples or send RAs or graduate students to NBAF-L to undertake the analyses. NBAF-W provides bioinformatic analyses, bioinformatic systems and software, as well as support and training and access to the popular Bio-Linux computing platform, optimized for bioinformatics research.

ANNUAL TARGETS AND PROGRESS TOWARDS THEM

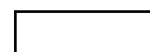
Capacity is defined by the availability of staff time, and all five nodes make $\geq 85\%$ of staff time available to users, with the remainder allocated to R&D. NERC has contracted to retain 50% of the print capacity at NBAF-L, and has bought into ~40% of the capillary capacity and 25% of the SOLEXA capacity at NBAF-E. 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. During this year there was again no significant equipment downtime.

SCORES AT LAST REVIEW (each out of 5)				Date of Last Review:	Jan 2007
Need	Uniqueness	Quality of Service	Quality of Science & Training	Average	
5.0	4.5	4.5	5.0	4.75	

CAPACITY of HOST ENTITY FUNDED by S&F %	Staff & Status: NBAF-S: T Burke (30%) DA Dawson (G7 100%), G Horsburgh (G7 100%), C Pagnier (G5 80%), Bioinformatician to replace A Frantz (G7 100%); NBAF-B: M Viant (G9 16%), U Sommer (G7 100%), J Byrne (G6 100%), Bioinformatician to replace L Bishop (G7 40%); NBAF-E: J Lovell (UOE7 50%), S Bridgett (UOE6 100%), A Gillies (UOE5 100%); NBAF-L: M Hughes (Res 8 50%) R Gregory (Res 7 50%), L Olohan (Res 7 100%), D Williams (Res 7 50%), L Parsons (Cler 5 20%); NBAF-W: D Field (10%), M Bicak (100%), T Booth (40%), tba Lead Bioinformatician (60%)	Next Review (March)	Contract Ends (31 March) 2014
		2012 (W), 2013 (B, E, L, S)	

FINANCIAL DETAILS: CURRENT FY						
Total Resource Allocation £k	Unit Cost £k			Capital Expend £k	Income £k	Full Cash Cost £k
	Unit 1	Unit 2	Unit 3			
FINANCIAL COMMITMENT (by year until end of current agreement) £k						
2010-11	2011-12	2012-13	2013-2014	2014-2015		

STEERING COMMITTEE	Independent Members	Meetings per annum	Other S&F Overseen
---------------------------	----------------------------	---------------------------	-------------------------------



NBAF SC	Chair (Prof M Ritchie) + 9	1–2	None (one SC for 5 nodes).
---------	----------------------------	-----	----------------------------

APPLICATIONS: DISTRIBUTION OF GRADES (current FY — 2010/11) – ALL NODES COMBINED

	$\alpha 5$	$\alpha 4$	$\alpha 3$	$\alpha 2$	$\alpha 1$	β	R*/Pilot	Reject
NERC Grant projects*	1	23	1	1	0	0	0	0
Other academic		2	2		0	0	1	0
Students	2 Pilot	24: 17 standard + 2 funded pilot + 5 not funded pilot	16: 6 funded + 10 pilot not funded	2	0	0	1	0
Pilot	1	11: 4 Funded + 7 Not funded	19 not funded	3	2	0	0	0
TOTAL [114]	4	60	38	8	2	0	2	0

APPLICATIONS: DISTRIBUTION OF GRADES (per annum average previous 3 financial years — 2007/2008, 2008/2009 & 2009/2010)

	$\alpha 5$	$\alpha 4$	$\alpha 3$	$\alpha 2$	$\alpha 1$	β	R*/Pilot	Reject
NERC Grant projects*	1	9	0	0.3	0	0	0	0.3
Other Academic	0	2	0	0.3	0	0	0.3	0
Students	0.6	14.6	9	2	0	0	2	1.3
Pilot	0.3	17.3	21.6	1	0.3	0.3	0	9.6
TOTAL [93]	1.6	43	30.6	3.6	0.3	0.3	2.3	11.3

PROJECTS COMPLETED (current FY – 2010/11) 132

	$\alpha 5$	$\alpha 4$	$\alpha 3$	$\alpha 2$	$\alpha 1$	β	R*/Pilot
NERC Grant projects*	3	40	0	0	0	0	0
Other Academic	0	3	0	0	0	0	0
Students	4	49	13	0	0	0	0
Pilot	1	19	0	0	0	0	0

Project Funding Type (current FY – 2010/11) (select one category for each project)

Grand Total	Infrastructure					PAYG				
	Supplement to NERC Grant *	PhD Students		NERC C/S	Other	NERC Grant*	PhD Student		NERC C/S	Other
		NERC	Other				NERC	Other		
132	8	20	15	2	2	33	19	13	4	16

Project Funding Type (per annum average previous 3 financial years - 2007/2008, 2008/2009 & 2009/2010)

Grand Total	Infrastructure					PAYG				
	Supplement to NERC Grant *	PhD Students		NERC C/S	Other	NERC Grant*	PhD Student		NERC C/S	Other
		NERC	Other				NERC	Other		
78.3	7.6	16	13	0.3	0.3	16.6	7.6	4.3	3.6	8.6

User type (current FY – 2010/11) (include each person named on application form)

Academic	NERC Centre/Survey	NERC Fellows	PhD Students	Commercial
250	20	12	74	2

User type (per annum average previous 3 financial years - 2007/2008, 2008/2009 & 2009/2010)

Academic	NERC Centre/Survey	NERC Fellows	PhD Students	Commercial
124	14	7.3	40.3	1

OUTPUT & PERFORMANCE MEASURES (current year)

Publications (by science area & type) (calendar year 2010)										
SBA	ES	MS	AS	TFS	EO	Polar	Grand Total	Refereed	Non-Ref/ Conf Proc	PhD Theses
0	0	14	0	95	0	0	109	82	14	13

Distribution of Projects (by science areas) (FY 2010/11)							
Grand Total	SBA	ES	MS	AS	TFS	EO	Polar
132	1	4	27.5	1	96	0	2.5

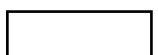
OUTPUT & PERFORMANCE MEASURES (per annum average previous 3 years)

Publications (by science area & type) (Calendar years 2007, 2008 & 2009)										
SBA	ES	MS	AS	TFS	EO	Polar	Grand Total	Refereed	Non-Ref/ Conf Proc	PhD Theses
0	0	9.6	0	55.3	0	1	66	51.6	3	11.3

Distribution of Projects (by science areas) (FY 2007/2008, 2008/2009 & 2009/2010)							
Grand Total	SBA	ES	MS	AS	TFS	EO	Polar
78.3	0	0.6	16.6	0	60.3	0	0.8

Distribution of Projects by NERC strategic priority (current FY 2010/11)

Grand Total	Climate System	Biodiversity	Earth System Science	Sustainable Use of Natural Resources	Natural Hazards	Environment, Pollution & Human Health	Technologies
132	8	97	3.5	5	1.5	12	5



OVERVIEW & ACTIVITIES IN FINANCIAL YEAR (2010/11):

This year was again the busiest so far experienced by the Facility; the recently opened nodes at NBAF-Birmingham and NBAF-Wallingford have built upon their first year's demand, with NBAF-B gearing up in its second year with 13 applications (10 approved; 3 NERC grants, 5 students). The existing facilities also saw a substantial further increase in the numbers of standard applications (45 to 57), numbers of users served (78 to 132 projects) and numbers of outputs (95 to 109). There was again a big response to the Small Grants Scheme (57 applications). Significant formal training was provided; at NBAF-S, two one-week courses were run for 6–12 participants at a time; NBAF-B ran the UK's first 3-day Environmental Metabolomics Masterclass, which included training in experimental design, analytical metabolomics and data analysis. NBAF-E ran a workshop on the new RAD sequencing method that can be implemented on the Illumina Solexa sequencer.

NBAF-B has implemented its first liquid chromatography mass spectrometry method, with further method development planned. Sequence capture methods have been introduced at NBAF-L using bespoke arrays to resequence targeted regions from non-model organisms, either on pooled population samples or on individuals, and the first projects have proved successful. At NBAF-E the major changes through the year to March 2011 have been the extension and upgrading of the Illumina sequencing platforms: NBAF-E now has access, through the GenePool Genomics Facility, to three GAIIX and two HiSeq2000 instruments, between them capable of delivering 70 billion bases of raw sequence data per day. A NERC research grant to develop the restriction site associated DNA sequencing has been started, and NBAF-E is collaborating with researchers across the UK to implement this 'game changing' technology.

SCIENCE HIGHLIGHTS:

Catching the Red Queen

Paterson S *et al* (2010) Antagonistic coevolution accelerates molecular evolution. *Nature*, **464**, 275-278. [*Impact factor 36.1*]

What drives evolution? One enduring theory is that the majority of evolutionary change is driven by an arms race between pairs of species, such as hosts and parasites, where an adaptation in one species leads to a counter-adaptation the other, and so on *ad infinitum*. This hypothesis is often known as the Red Queen Hypothesis, since, as the Red Queen explains to Alice in Lewis Carroll's *Through the Looking Glass*, "here, it takes all the running you can do, to keep in the same place." Using next-generation sequencing, Paterson *et al* were able to follow the emergence and rise of mutations within experimental populations of viruses and their bacterial hosts and, for the first time, test this important theory experimentally. They showed that the genome sequences of viruses running with the Red Queen evolved twice as quickly as those that evolved towards a fixed target. Furthermore, while evolution towards a fixed target led to a set of similar-looking virus sequences, Red Queen evolution led to increasingly divergent virus sequences. These findings suggest that interactions between species drive the majority of evolution and that by causing rapid divergence they could potentially lead to speciation itself. NBAF-L provided sequencing and assistance with bioinformatic analysis.

Natural and sexual selection in a wild insect population

Rodríguez-Muñoz *et al.* (2010) *Science*, **328**, 1269. [*Impact factor 31.4*]



The understanding of natural and sexual selection requires both field and laboratory studies to exploit the advantages and avoid the disadvantages of each approach. However, studies have tended to be polarized among the types of organisms studied, with vertebrates studied in the field and invertebrates in the lab. This study combined video monitoring with DNA profiling of all of the members of a wild population of field crickets across two generations to capture the factors predicting the reproductive success of males and females. The factors that predicted a male's success in gaining mates differed from those that predicted how many offspring he had. The study confirmed the fundamental prediction that males vary more in their reproductive success than females. Females as well as males left more offspring when they mated with more partners. NBAF-S supported the microsatellite marker development and genotyping.

An aerial-hawking bat uses stealth echolocation to counter moth hearing

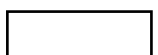
Goerlitz *et al.* (2010) *Current Biology*, **20**, 1568-1572. [*Impact Factor 10.0*]

Ears evolved in many nocturnal insects, including some moths, to detect bat echolocation calls and evade capture. Although there is evidence that some bats emit echolocation calls that are inconspicuous to eared moths, it is difficult to determine whether this was an adaptation to moth hearing or originally evolved for a different purpose. Aerial-hawking bats generally emit high-amplitude echolocation calls to maximize detection range. This study presented the first example of an echolocation counterstrategy to overcome prey hearing at the cost of reduced detection distance. The study combined comparative bat flight-path tracking and moth neurophysiology with faecal DNA analysis to show that the barbastelle bat emits calls that are 10 to 100 times lower in amplitude than those of other aerial-hawking bats, remains undetected by moths until close, and captures mainly eared moths. Model calculations demonstrated that only bats emitting such low-amplitude calls hear moth echoes without their calls being conspicuous to moths. This stealth echolocation allows the barbastelle to exploit food resources that are less available to aerial-hawking bats emitting calls of greater amplitude. NBAF-E and NBAF-S supported the sequencing and genetic diet analysis.

Giant virus in marine food webs

Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010) Giant virus with a remarkable complement of genes infects marine zooplankton. *PNAS*, **107**, 19508-19513. [*Impact factor 9.8*]

Zooplankton feed on bacteria and phytoplankton and play a major part in the functioning of marine food webs. Such marine food



webs provide vital ecosystem services through nutrient cycling and CO₂ absorption, and help support fisheries, but the factors that affect the abundance of different components of these food webs are poorly understood. It has recently been discovered that viruses of zooplankton are highly abundant in seawater and so have the potential to regulate zooplankton abundance and ecosystem function. Remarkably, these viruses are amongst the largest known of any viruses and have a correspondingly high number of genes. Here, Fischer *et al* performed a genetic analysis of the largest of these viruses and discovered more than 500 genes within its 730-kb genome. They then developed microarrays to dissect the expression of these genes throughout the process of viral infection. They found that these viruses encode a range of metabolic activities normally encoded by their hosts, and their results shed light on the evolution of large genome size in these viruses. This was an international collaborative project, in which NBAF-L provided microarray analyses and expertise.

Cultural inheritance drives site fidelity and migratory connectivity in a long distance migrant

Harrison *et al.* (2010) *Molecular Ecology*, **19**, 5484–5496. [*Impact Factor 6.5*]

Cultural transmission is thought to be a mechanism through which migratory animals exploit new habitats, but little evidence exists in wild populations because of the difficulty of following individuals over successive generations and wide geographical distances. Cultural inheritance of migration routes represents a mechanism whereby geographical isolation can arise between separate groups and could constrain individuals to potentially suboptimal sites within their range. Conversely, adopting the parental migratory route in adult life, rather than dispersing randomly, may increase an individual's reproductive success because that strategy has already been proven to allow successful breeding. This study combined a pedigree of related light-bellied Brent geese (*Branta bernicla hrota*) with 6 years of observations of marked birds to calculate the dispersal distances of adult offspring from their parents in both Ireland and Iceland. In both countries, the majority of offspring were found to recruit into or near their parental sites, indicating migratory connectivity in the flyway. Despite this kin structure, there was no evidence of genetic differentiation using genotypic data from 1127 individuals across 15 microsatellite loci. The study suggests that the most migratory connectivity among subpopulations is far more common than previously expected. The results are significant to the future conservation and management of goose populations. NBAF-S supported the microsatellite marker development and genotyping.



Lohse K, Sharanowski B, Stone GN (2010) Quantifying the Pleistocene history of the oak gall parasitoid *Cecidostiba fungosa* using twenty intron loci. *Evolution*, **64**, 2664–2681. [*Impact Factor 5.659*]

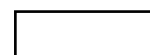
The longitudinal spread of temperate organisms into refugial populations in Southern Europe is generally assumed to predate the last interglacial. However, few studies have attempted to quantify this process in nonmodel organisms using explicit models and multilocus data. We used sequence data for 20 intron-spanning loci (12 kb per individual) to resolve the history of refugial populations of a widespread western Palaearctic oak gall parasitoid *Cecidostiba fungosa* (Pteromalidae). Using maximum-likelihood and Bayesian methods we assess alternative population tree topologies and estimate divergence times and ancestral population sizes under a model of divergence among three refugia (Middle East, Balkans and Iberia). Both methods support an “Out of the East” history for *C. fungosa*, matching the pattern previously inferred for their gallwasp hosts. However, coalescent-based estimates of the ages of population divides are much more recent (coinciding with the Eemian interglacial) than nodal ages of single-gene trees for *C. fungosa* and other species. We also find that increasing the sample size from one haploid sequence per refugial population to three only marginally improves parameter estimates. Our results suggest that there is significant information in the minimal samples currently analyzable with maximum-likelihood methods, and that similar methods could be applied to multiple species to test alternative models of assemblage evolution. How will ecosystems change under anthropogenic climate change? This study begins to address the complex question of how ecological communities reassemble after major climatic change. Using multiple loci gives much greater resolution in space and time, and this work starts to define new paradigms in multi-locus (and moving to whole-genome) based analyses. NBAF-E provided advice on experimental design and performed sequencing.

FUTURE DEVELOPMENTS/STRATEGIC FORWARD LOOK

Ultra-high-Throughput Sequencing

2010–11 saw an increase in both the number and diversity of next-gen sequencing projects conducted at NBAF-L and NBAF-E. These projects include de novo sequencing of bacteria and transcriptomes, RAD tagging, genome resequencing, sequence capture, meta-genomics and digital gene expression. With advances in sequencing technologies we may expect these applications in the near future to be extended to include de novo eukaryote sequencing and epigenetics (bisulphite sequencing and ChIP-Seq). Both NBAF-L and NBAF-E form part of larger sequencing facilities – The Centre for Genomic Research (CGR) at Liverpool and GenePool at Edinburgh – that both also receive core funding as MRC hubs, and 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.

NBAF-L has recently upgraded its 454 platforms to a 'Phase E' longer read system, which will provide researchers with significant advantages for amplicon sequencing and de novo assembly of environmentally relevant organisms. NBAF-L will shortly upgrade all three of its SOLiD platforms to the 5500xl system, which will provide up to 300Gb/run or 30Gb/day for resequencing and gene expression studies. NBAF-L benefits from early access to these technologies. NBAF-L has also taken delivery of an IonTorrent Personal Genome Sequencer, which has the potential to offer extremely rapid and cheap sequencing for diagnostic and metagenomic projects. Delivery of a 454 GS-Junior sequencer is planned, which may be ideal for smaller de novo transcriptome or amplicon sequencing projects. NBAF-L has also purchased an Illumina GAI sequencer, although Illumina sequencing for NERC users will



continue to be provided via NBAF-E rather than NBAF-L.

NBAF-E has installed the latest instance of Illumina's platforms, the HiSeq2000. The facility now has three GAIIX and two HiSeq2000. The HiSeq2000 delivers an astonishing 200 Gbase of data per 8-10 day run, and is already being used for genomic, genetic and transcriptomic NERC science. The GAIIX instruments, though they have each only ~one-third the throughput of a HiSeq are still useful for smaller projects, or projects requiring longer reads (only the GAIIX can currently generate 150-base reads). NBAF-E now also has access to improved mate-pair technology, based on in-house development and collaboration with the Sanger Institute: this is essential for contiguation of large genomes. The future roadmap of the HiSeq is to deliver 600 Gbase/run (or 80 Gbase/day), and this upgrade is in the process of installation in Edinburgh. The increased throughput of the instruments requires upgrading of the throughput of the sample preparation laboratory (new robotic systems, and automated library preparation systems), and a rapid growth in the bioinformatics capacity of the Facility (including 100 Tbyte local storage, >150 new compute cores, and a large RAM analytical instrument for genome assembly). The facility's Roche 454 instrument is being upgraded to the 'PhaseE' longer reads.

Both nodes are investing heavily in shared, compatible laboratory management software to track all aspects of projects and we hope to use these to provide greater feedback to users on the status of their projects. Both nodes are also investing in computing infrastructure to cope with assembly, mapping and archiving of increasing volumes of data. Finally, the centre managers for NBAF-L and NBAF-E are working closely to ensure that users are directed the most appropriate node for their requirements.

Both the Edinburgh and Liverpool host laboratories have received additional core funding from the MRC, doubling their size and quadrupling their output, which started in mid-2009. This offers NERC users access to ever more sophisticated capability in pursuit of NERC science goals.

NBAF is also watching the development of third-gen technologies (Pacific BioSciences, Visigen and Nanopore Technologies), which are likely to start to deliver in the next 6-12 months. These single-molecule technologies promise direct sequencing without amplification bias, and if the manufacturers achieve the necessary parallelisation, then throughput could be gigabases *per hour*, with delivery of high-quality long (kilobase plus) reads. These instruments will be of even greater interest to NERC science, where often the materials are rare and valuable. True population genomics will become feasible, and thousands of whole genomes might be generated at minimal cost. Again, continued investment in both instrumentation and bioinformatics capacity is essential to grasp this opportunity, but NBAF is well placed to lead on introducing these technologies.

Gene expression profiling

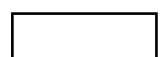
Microarrays continue to provide the most straightforward route to expression profiling and genome-wide genotyping. Probe density continues to increase, costs decline, and flexibility and speed of fabrication are rapidly improving. Recent projects have demonstrated how this new capability can reduce the time and cost required for full-blooded expression profile experiments, even for non-model species. These include projects where array design can be integrated with de novo transcriptome sequencing. This enormous productivity increase substantially changes all previous expectations and capability. Several NERC projects are now exploiting digital gene expression methods (RNAseq and SAGE) and we expect this trend to continue. NBAF-L is well placed to service this need since the CGR at Liverpool is a leading facility in the UK for digital gene expression, and has invested heavily in robotics for library preparation and bioinformatic and statistical methods for analysis.

Statistical analysis of array data has also advanced significantly recently. Thus, the combination of more comprehensive ANOVA-based loop designs, of greater levels of independent replication, and the use of improved analysis of false discovery rate criteria for calling genes as differentially expressed (DE) have all increased sensitivity of the technique and the delineation of responding metabolic pathways and biological processes. Moreover, commercial network-based tools (e.g. Ingenuity) for the discovery of responding gene networks are now fully deployed. With its new NERC-based statistical provision, NBAF-L has also developed a rank-mean method for testing the properties of gene sets, GO categories or other gene pathway lists, making it possible to detect much more subtle effects at the pathway level than before.

The landscape of transcriptome analysis is changing rapidly, with counting-by-sequencing methods (RNA-Seq and DeepSAGE) promising to replace and supercede the traditional microarray platforms. These new methods have great promise for NERC science as they do not rely on a well-analysed genome or transcriptome to permit analyses (NERC science organisms are often 'neglected' and lack well developed genomic resources), and they permit the discovery and definition of novelty (i.e. they are not limited to the reporting of what is 'known'). Both NBAF-E and NBAF-L offer RNA-Seq and DeepSAGE for NERC science, and NBAF-E are continuing to develop in-house protocols for DeepSAGE in particular.

Population Genomics

SNP typing on the Illumina BeadXpress was introduced last year and several local/ R&D projects have now been completed, with results that improved each time and are now of the highest quality. There were some initial issues around SNP calling and DNA quality that have been fully resolved. A budget is available to support 2-3 small-scale PhD projects in the coming year, to make this technology more widely available as an alternative to microsatellite analysis. There are significant timesavings to the investigator but the system does require more dedicated technical support than microsatellite typing. It is unclear whether SNP typing will supplant microsatellite analyses, as demand for microsatellites remains very strong. This has been encouraged in birds in particular by our development of conserved primers that amplify polymorphic loci in a wide range of species. We have recently successfully trialled the use of 454 sequencing in microsatellite development and introducing this approach should reduce the sequencing costs and improve the results obtained with that technology. We introduced reference-strand conformation analysis (RSCA) as a new technique 1.5 years ago and demand for this in the analysis of MHC genotypes has been strong in the last year, with several projects successfully completed. This approach may eventually be supplanted by new-gen sequencing methods but this relies on tagging each individual, which is currently costly and informatically complex.



With the rapid drops in the cost of raw sequence generation, direct sequencing approaches are becoming more and more attractive for population genomic analyses. An important application is that of reduced representation sequencing, where a specific and repeatable portion of the genome is selected for sequencing, and SNP and other variants called and scored in this smaller, but representative, span of the genome. Reduced representation can be achieved through targeted sequencing, where a selected known portion of the genome is specifically extracted from the whole (using array-based or in-solution oligonucleotide hybridisation and pull-down), and (re)sequenced to depth. Both NBAF-L and NBAF-E have experience of using this targeted resequencing technology in both model and non-model organisms, and it is a very promising avenue for many genomes of ecological interest. A second reduced representation technology is Restriction site Associated DNA sequencing (RAD), where sequencing is limited to the portions of the genome adjacent to the sites of cutting of selected restriction enzymes. NBAF-E has pioneered this technology in the UK (and Europe) and RAD-Seq is one of the most popular offerings now in the next-generation NBAF 'menu'. In addition to developing new wet lab techniques for RAD-Seq, Edinburgh has been developing analytical tools for RAD data and promoting the technology by hosting a UK user group and workshop.

Metabolomics

Sample preparation: in response to community demand, NBAF-B will develop optimised standard operating protocols for the rapid sampling and quenching of microbes and algae from freshwater and marine samples, coupled with efficient metabolite extraction.

LC-MS of polar metabolites: building on the recent development of a HILIC-based LC-MS method for targeted metabolomics (as part of project NBAF-435), NBAF-B now plans to implement a 'shotgun' (i.e. discovery based) LC-MS metabolomics approach based around the same HILIC chromatography method. This will provide a useful complement to our established direct infusion MS methodology. In addition, the development of further targeted LC-MS metabolite analyses is planned on the new TSQ Vantage triple quadrupole mass spectrometer.

Metabolite identification: NBAF-B staff are now completing the validation of a novel software tool for the identification and quantification of metabolites in 2D J-resolved NMR spectra, the mainstay NMR technique offered by the facility. If successful, we plan to implement this new software tool into routine operation by NBAF-B users.

Environmental lipidomics: Technical challenges remain in the global measurement of lipids, including the need for improved bioanalytical and bioinformatic tools. The NBAF-B direct infusion MS method has been implemented for lipidomics (as part of projects NBAF458 & NBAF459). Development will now be focused on establishing an LC-MS setup to resolve lipids of the same class on a reversed-phase chromatography column.

Bioinformatics

Next-generation sequencing has vastly increased throughput per instrument. Meeting the demand for processing, storage and data handling is a widely recognised problem. In large part, NBAF-L and NBAF-E are well-placed to deal with these issues, given their extensive expertise in sequence generation and analysis across a wide variety of human, agricultural and environmental projects, and given their investment in computing infrastructure. Nevertheless, NERC projects, on non-model or environmental samples, can present unique challenges and the ability to react flexibly to researcher needs via the provision of bespoke scripting/programming services is essential to the NERC user community. This scripting service is provided by NBAF-W. Scripts developed are made available to the wider community *via* the Analysis Scripts section of the NEBC website. NBAF-W also adds value by providing increased access and interaction with data to researchers through standard or bespoke databases and interfaces, with services in data modelling and database implementation. Good data organisation facilities like these are often core to effective analysis and, as datasets grow in size, this facility is likely to become increasingly important for NERC researchers.

NBAF-W also provides access to new analysis tools, especially those applicable to new data types through the Bio-Linux system, its packages and the Bio-Linux Bioinformatics Documentation Project. For those applying through NERC grants, NBAF-W will also engage in data management activities, including the NEBC Handlebar sample management system, data cataloguing in the NEBC Envbase data catalogue, support for data annotation that meets approved standards, and support in submission to public archives such as EMBL and Genbank (nucleotide sequences), ERA (next-generation nucleotide sequences), ArrayExpress (microarray data) and the BioInvestigation Index (project-level metadata).

NBAF-E has continued to develop data analysis tools for the NERC community through collaboration with bioinformatics researchers. In particular, 2010 saw the release of RADtools, a suite of programmes for the processing and analysis of RAD-Seq data, jMOTU, a user-friendly tool for the analysis of metagenetic data (such as Roche 454-sequenced amplicon pools from microbial diversity research), and iPhy, a phylogenomics workbench that simplifies the tasks underpinning the collation and analysis of large-scale phylogenetic data. These sorts of developments, embedding best-practice tools in easy-to-use interfaces, are set to be a key future goal of NBAF, as these tools are essential for the dissemination of expertise in next-generation data.

