Community Network in metagenomics (ComMet) Inaugural meeting

The exploitation of metagenomics and meta-omics approaches in life science research

Hosted at the School of Life Sciences, University of Warwick

Tue 15th July Closed session for committee only

Discussion of network strategy

Present at committee meeting

Prof Liz Wellington (EMH)Uni of Warwick
Dr Emma Travis (ET) Uni of Warwick
Prof Conrad Bessant (CB)Queen Mary's London
Dawn Field (DF) University of Oxford
Jack Jordan (JJ) with views and comments from Dr John Ward Imperial College London
Penny Hirsch (PH)Rothamsted Research
Dr Rob Finn (RF) EBI
Dr Julian Marchesi (JM) Uni of Cardiff/UCL
Dr Dieter Bulach (DB) Monash University

Apologies received from

Dr Richard Leggett TGAC
Dr Alan Walker University of Aberdeen
Michael Ball (BBSRC)

Prof Liz Wellington welcome and presentation of the network (link to presentation)

Aims and objectives round table discussion on strategy:

- Overall strategic aims to maximise impact of network on UK meta-omic productivity
    - We need active community
    - Priority to promote links between people from academia and industry across a diverse range of disciplines
    - Important to increase industrial involvement
    -
- Annual workshops
    - discussed how best to deliver the program – 3 main workshops and also one day meetings? London, Edinburgh, Birmingham?
    - Host across the UK - One day meetings in London to reach wider audience/ increase accessibility
    - Each meeting to have specific theme, focus, good speakers and discussion

    - Idea of 1 day meetings to focus on a particular theme ie 2 day microbiome meeting in London; specific one to appeal to industry. Market it as industry chance to work with leading PIs
    - Suggested London for applied microbiomics –human plant, animal, gut colicins, antimicrobials,(biocontrol away from vaccines) skin enzymatic interactions, micorbiome of worm and sponges, probiotics
    - JM Action – put together a draft programme and come up with some dates (Jan/Feb). Fund invited speakers others have to find funding
    - Workshop at EBI in at least a year. Rob needs to talk to training team 3 day workshop 40 is max no. approx. £300 per head cost
        - If embrace other portals as well then longer course (IMG MG-RAST)
    - Have a workshop on metaproteomics/functional side at the end of the grant

- Staff and student exchanges and knowledge transfers – how to structure?
    - Exchange of PDRAs and PhDs to other labs in UK/internationally
    - Have funding available on website and how to apply
    - Funding committee to award funds
    - Exchanges fo PDRAs and PhDs depends on how funding split how many
    - On website a section on mobility – people can find and apply to exchange with labs with significant expertise. Make a case. Doesn't have to be UK based (check BBSRC)
    - something specific on analysis side maybe a TGAC-EBI link to develop a training programme for 3 of best students to learn to sequence and analyse
- Website development (including editable section to run akin to wiki), forum, blog, social media and other online resources
    - develop website and e-learning environment – post talks on there
    - ET presented existing website run from the University of Warwick and twitter account
    - Felt strongly that it needed to be independent website eg metagenomics.org.uk (RF purchased) or metagenomics.uk
    - Action: ET to determine if can be hosted through Warwick or find alternative host
      Need identity/ logo. Discussed ideas and RF to implement logo of a wire sphere with comet tail /bacterium comet
    - Rather than have additional forum, link to exisiting forums /communities existing eg seqanswers and biostar. Decided for forum use Seqanswers metagenomics and direct people to the existing resource. Have explanations for potential users of forum.
    - Have closed section for members only
    - information portal for academics and scientists not directly in field
    - Implementation of e-learning environments
    - Online learning tools, tutorials and pipelines.
    - Encourage other UK academics to get onboard to analyse datasets
    - Don't want to duplicate EBI portal eg example datasets but make it more biological
    - Databases to be made available for sharing noted that EBI have 2 years lag and then made public
    - Fosmid/BAC libraries are storable and could be shared/distributed
    - Cumulative resources
    - Training resources including seminal papers and reviews – make this more international eg Janet in Antarctica. Put presentations from overseas on.
    - Want to be referring people to existing resources

- Online tools
    - Guidance as to where to go
    - More of a biological approach
    - Alternative to chime
    - Page on analysing taxonomic structure
    - Protocols for potential users
- Community helpdesk would be good
- Illustrate stories of what is happening - Make accessible to general public with videos showing metagenomics
    - Link to probiotics microbiome outreach
    - journey to centre of the earth PH
    - NERC podcast on science
    - TGAC sequenced Chris Packham's garden
    - EMH talking at botanical gardens
    - EBI video talks
    - Encourage members to go to a place to do a podcast to make visually appealing
    - The health services do lots of little videos
    - Embarrassing bodies JM
    - animal microbiomes eg sheep feet DB
- Encourage involvement from across disciplines to contribute to e-learning resources – run through the website. One tab for resources/outreach. Members of the community to add to this. ET responsible for consistency
- Have funding calls and funding committee need website page for funding with BBSRC logo on
- Embed google analytics to log hits
- Website page about grants and activities
- Want to attract people with attractive funding
- Offering grants for funding for meetings/ build credit for contribution to workshops by contributing to website/e-learning
- Need to expand the community and have page so that people know who is involved
- Disciplines identified: Food microbiology and disease; Animal health; Human health; Synthetic biology; Understanding metabolism; Combating drug resistance

- Capacity building strategies
    - Identified hurdles in field –discussed portals/resources and differences between them. IMG, EBI, MG-RAST. Use network to determine preferences. Design protocol to transfer data between platforms?
    - Metadata essential
    - metagenomics standards SIG journal DF
    - Building awareness in other disciplines; interdisciplinary collaborations
        - Write comments for journals of relevance to raise awareness
        - Micro and macro-organisms need to be represented – fungi, plant roots, seeds
        - Host microbiome is big growth area need more representatives in this area
        - Increase openness of meetings to reach wider audience
    - Links with industry
        - Noted a number of existing links to a wide range of industries that may have an interest
        - Prioritise next workshop to encourage industry awareness

        - Action: promote awareness with contacts in industry
            - John Ward has lots of industrial contacts Novartis, GSK
            - Exiting contacts –EBI industry/ BBSRC contacts/ biotechnology/ probiotic companies (Yakult/S.Wales/Netherlands/Devon)
            - Invite L'Oreal, Unilever
            - On the plant side – PH contacts
            - JM Synthetic biology economy of metabolism to model microbial metabolism, economics models
            - Syngenta
            - AstraZeneca
            - Plant/soil microbiomes
            - Natural product discovery for herbicide/insecticides
            - Novozyme –EMH (plants in rhizosphere)
            - Unilever (CB, EMH)
    - Outreach activities resource to schools and general public – local schools; big bang fair; social media
        - ET discussed existing activities in local primary school
        - Provide resources for outreach activities through website

Wed 16th July - Open session ***(Room BSR3)***

The format of the open sessions will be a series of vignette talks (25 mins) where participants can present current research in the field of metagenomics, in particular highlighting problems encountered. This will be followed by discussion sessions. Facilities for breakout sessions are available.

Morning session –A Data Collection

9.30-12.15

9.30 Welcome Prof Liz Wellington

To include discussion of sample collection, processing, high-throughput sequencing, depth of sequencing and level of replication needed

9.40 Dr Nick Loman (University of Birmingham) "Long read technologies for metagenomics"

Nick presented a great talk on the possibilities that long read technologies offer in metagenomics. His alternative talk title of 'Stop the madness' provided acknowledged the difficulties associated with assembling short reads in a meaningful manner. The main questions are typically: who is there?; what can they do?; who does what? The idea with the long read technologies is that one can look at strain level variation. The read needs to be long enough to cover the common repeat elements, for instance >10kb in the case of PacBio. The three long read technologies discussed were Moleculo (Illumina,) PacBio, and Oxford Nanopore. For Moleculo (Illumina) library preparation involves fragmenting the DNA to about 10kb. Samples are split into 384 pools, which, in addition to standard illumine adapters have a tag for each pool. Sequence analysis involves demulitplex into pools and use as synthetic long reads prior to assembly. Pacific biosciences (PacBio) is in use at Liverpool, TGAC and Sanger, Oslo and Yale. It works well for genome assembly and turns noisy error rate prone sequences into accurate sequence. It requires enough sequences to obtain consensus. It first aligns longer reads then matches shorter reads in an iterative assembly. PacBio can detect base modifications through looking at the largest contigs. This makes it potentially useful for epigenetics. Nick presented Oxford Nanopore as the third option, an exciting and very novel approach which he is beta testing. It is an USB sequencer and flow cell which is a portable and single molecule sequencer. It works though slotting a nanopore into biological membrane and apply current. With molecules in the nanopore there is a change in the electrical signal. It could be used as a general use biological sensor. For DNA can get 4 5 or 6 bases in contact and can detect sequence, then get it feed at constant rate. Library prep involves genomic DNA prep. Need about 1 ug and an end repair adapter hair pin. Sequence both sides to reduce error. Need adapter and motor attaching. Excitingly the read length is limited by size you put in with no effect of read length on error quality allowing up to 50kb to be read. It is error prone. And there is a need to optimise forward and reverse sequences to minimise error. As it is streaming software it is possible to have real time data.

10. 10 Dr Penny Hirsch and Dr Ian Clark (Rothamsted Research) "Sampling the soil metagenome and metatranscriptome"

Penny Hirsch's talk illustrated the importance of the samples chosen and the metadata collected. Rothamsted research is home to the oldest, continuous agronomic experiments in the world, with vast amounts of metadata available, an invaluable resource. Metagenomic libraries have been constructed from soil under different land use and microbial communities and functions compared.

10.40 Coffee ***(BSR3)***

11.10 Dr Daniel Morais (Brazilian Microbiome Project/Rothamsted Research) "Brazilian Microbiome Project: revealing the Unexplored Microbial Diversity - Challenges and Prospects"

Daniel Morais gave an interesting insight into a Brazilian network which is drawing researchers interested in microbiomes together. Brazil has having 20% of biological diversity in terms of macro organism and became a member of Global International Biodiversity. He highlighted some of the issues in the field, such as lack of standardisation, replication and poor experimental design. The network is working on standardised protocols and encouraging groups to talk and compare data. Ion Torrent and MiSeq are the most common platforms used and efforts to compare datasets and make them comparable are underway.

11.40 Dr Dieter Bulach (University of Monash)

Continuing with the international theme, Dieter Bulach provided a valuable insight into his work in genomics and metagenomics in Monash, Australia. Some of the issues to be overcome were metadata quality, access to primary data and the importance of quality primary data, annotation and out of date annotation. He highlighted the importance of asking specific questions when looking at whole metagenome data and illustrated his points with a study of faecal microbiomes to investigate the impact of addition of probiotics to preterm babies for protection against necrotising enterococci.

12.15-13.15 Lunch ***(BSR4)***

Afternoon session –B Data analysis

To include discussion of current pipelines, databases, online tools and data accessibility

13.15 Professor Conrad Bessant (Queen Mary University London) "Metaproteogenomics in Galaxy"

Conrad Bessant gave an interesting presentation exploring the potential to use the Galaxy platform with a web-based interface for analysis of meta-omic data. He showed how different tools could be linked up within Galaxy, enabling integration of customisable tools. This can allow reasonably comprehensive transcriptomic and proteomic profiling with no prior knowledge.

13.45 Dr Julian Marchesi (School of Biosciences, Cardiff University)

Julian Marchesi provided a very different way of looking at metagenomics, through functional metagenomics, where DNA is put in the host and expressed. Whilst it is hard to do, Julian believes this is the only route for novelty, to provide novel building blocks for synthetic biology. Extracted DNA is inserted into a PUC vector, then high throughput robots are employed to pick colonies. Selective media allows identification of clones of interest which can be investigated further with transposon mutagenesis knockouts to find genes, or accessory genes of interest. Julian discussed a variety of screens which could be used to identify the novel genes of interest and talked about the use of the gateway system to allow insertion of DNA of interest into different hosts, which can be essential for expression of the gene of interest.

14.15 Dr Rob Finn and Dr Alex Mitchell (EMBL-EBI) "Metagenomics analysis services at the EBI"

Rob Finn and Alex Mitchell provided valuable insights into the great resource to the meta-omic community provided by EBI – the EBI metagenomics portal. This has been running for 3 years and provides an integrated European nucleotide archive for DNA submission, data analysis and presentation of data. DNA is

uploaded using sequence loader. Templates and check lists are used to add metadata. The importance of the metadata was highlighted, the who, where, when, what and how of the data. Importantly, the alignment at EBI is not against reference sequences, rather it uses the reads. Therefore the models are generated from multisequence alignment rather than homology. A huge amount of data is submitted to EBI, who have had the challenge of scaling to accommodate this. Rob highlighted the flexibility in the pipeline and the integration of new tools. He also talked about the importance of re-analysis of data as annotation keeps improving.

14.45 Dr Andrew Millard (Warwick Medical School, University of Warwick) Title TBC

Andy Millard presented an insight into approaches he has applied to viral metagenomes. Metaflow kraken and web based kraken are very fast for RNA analysis. Issues with analysis of viral genomes arise from the diversity of different types of virus which need different extraction methods. There are no universal methods and a range of molecular methods. Andy raised the issue of the Camera database closing down and also highlighted that Metavir2 is a user friendly web interface for looking at viruses.

15.30 16.00 Tea ***(BSR3)***

16.00 Dr Christopher Quince (Warwick Medical School, University of Warwick) "Extracting species and strain-level variation from metagenomics data"

Chris Quince presented novel techniques for extracting species and strain level variation from metagenomics to look at the vital questions: 'Who is there?, what are they doing? And who is doing what?'. He discussed the problems involved in assembling genomes from metagenomic data. The analysis of community population genomics requires strain level variation. He presented CONCOCT, a new algorithm that combines sequence composition and coverage across multiple samples, to automatically cluster contigs into genomes.

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, and Quince C Binning metagenomic contigs by coverage and composition. Nat Methods 2014 Nov;11(11):1144-6

16.45-17.30 Sum up and define actions discussion led by Liz Wellington and Chris Quince

8pm Dinner at ***Radcliffe House*** private dining room (University of Warwick)

Thursday 17th July - Open session ***(BSR5)***

Morning session-C Molecular screening tools and techniques

9-10.30

To include development of novel screening tools eg coupling novel transcriptional regulators and reporter gene assays

Dr Andrew Neal and Dr Ian Clark (Rothamsted Research) "An assembly-free, gene-centric approach to studying functional redundancy in soil microbial communities"

Andy Neal presented a fascinating talk on using metagenomics to establish how best to manage land as well as possible. He explored issues of resilience, functional redundancy and the need to understand at a system and process specific scale. Through designing a pipeline in Galaxy and using a curated set of amino acid sequences for specific proteins, Andy has investigated phytases, key enzymes for ensuring phosphorus is accessible for plants. He found that there is a more capable community in the fallow soil.

Jack Jeffries (UCL) "Sequence led Functional Metagenomics"

Jack Jeffries (from Prof John Ward's lab) presented a very different approach. In single genomic mining, libraries are created and then functions are identified, the proteins responsible identified and thus the genes are found by comparing with the genome sequence. Jack presented some work on tongue microbiomes from healthy and diseased mouths to demonstrate how this approach could be used for metagenomic approaches. It is a very focussed approach and can be of interest to industry. The processes involved need to be very high throughput.

10.30 -11.00 Coffee ***(BSR5)***

Morning session D Expression screening techniques/hosts

11-12

To include sensitivity of screening methods for detection of low level gene expression, enzyme activity-based functional screening and developing small molecule-based functional metagenomic screening. To include discussion of a variety of hosts for functional metagenomics, including Gram positive, Gram negative bacteria as well as Archaea; using host engineering to improve performance of commonly used hosts:

Dr Wei Huang (Department of Civil and Structural Engineering, University of Sheffield)

Wei Huang presented a fascinating, again with quite a different perspective to previous talks. His work combines stable isotope probing (SIP) to narrow down the population of interest to then investigate with metagenomics. He also used single cell Raman technology mediated metagenomics, where the cells of interest can be investigated by laser induced ejection. Through a clever integration of biosensor technology and SIP organisms and genes of interest can be identified.
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0047530

Dr Gregory Amos (School of Life Sciences, University of Warwick)

Greg highlighted the urgent need to discover new antimicrobial drugs due to the constant emergence of bacteria resistant to multiple antibiotics. Metagenomics allows the exploitation of the unculturable

bacterial fraction present in the environment. Using libraries from a variety of environments, metagenomic techniques were applied to explore the secondary metabolite biodiversity, with novel gene clusters identified and characterised.

12.15 pm lunch ***(BSR4)***

Afternoon session –Break out session 4 groups plus rapporteurs to respond to specific questions concerning delivery of network objectives

1.15-3.00 Break-out sessions

3.00 Tea ***(BSR3)***

3.15-4.00 Report sessions and wrap up

Comments from break out groups:

- Network to be a diverse group to bring people together, encourage communication and alternative ways of doing things
- Cross fertilisation of ideas
- Lots of desire to work together within the metagenomics community. There is an element of being joined up internationally.
- Identification of members on the website via expertise - using tag clouds –who does what/ has expertise/ has applied metagenomics on projects/
- Databases
- KeGG is a useful way-in but unstable
- Is there somewhere to put algorithms
- Networks - using people and the cloud to build pathways
- Depositing PFAM for individual processes/pathways
- Drive traffic to metagenomics portal (EBI) - more datasets will allow cross-correlation across datasets - analysis of 'dark matter'. Signal from the noise.
- Identification of vulnerable databases
- Metadata standardises
- Annotation improved
- Dark matter
- Input into synthetic biology
- How to use metagenomic data to serve industry
- Big industry questions
- Some labels metagenomics/ some population genomics
- Create funding opportunities
- Scope for metagenomic centre
- Training- e-learning- who can edit? How best to present/ promote?
- Link to SGM – attract students and funding?
- Communicate back to BBSRC
- Blogs on papers published link and follow blogs/tweets.
- The network as a source of information and best practice in deposition. Is there a place to put that information?
- Value of databases are in the updating - pipelines need to be updated in response to the latest tools.

- Crowd sourcing routes to curation - Manchester Computer Science doing text-mined routes to annotation.
- Crowd sourced cycles and pathways - wiki pathways already exists. A group on wiki pathways?
- Network - disseminate curated network sets as much as possible
- From the point of view of EBI - as a service provider, requires feedback
- Pipeline testing, go through add links to curated data - will then add to Galaxy toolshed
- Pin-point outputs and where tools should be
- Pin point unmet needs and who should action them
- Comparison of datasets - conclusions through comparisons
- Have carefully thought through questions
- Development of software to answer questions
- Differences between environmental and medical - in number of replicants
- Lots of diversity between different replicants
- Need for training, how to analyse data - from an unbiased environment
- Accessability
- Blogs and tweets
- Integration of things into resources
- Info about papers - self-fulfilment. Blogs better than tweets.
- Reciprocally link or follow tweets.
- Announcements in journals
- Ideal metagenomics experiments, as a point of reference - with realistic information regarding resources required. Costs vary
- Information for those new to field to introduce it.
- Outreach to schools/ general public/ scientists/industry
- Link the website to seq answers for a forum
- Page on website on who provides services

Michael Ball also provided us with a helpful insight into the drive for the network from BBSRC. It is recognised that metagenomics research happens in different areas (sequence vs frunctionsal; environmental vs medical; functional vs assembly). Wanted to bring together people to share best practises, methods/ tools and identify where there are knowledge gaps. BBSRC can also use it as a tool to share funding opportunities. It will increase information flow between groups.