Exploratory Conference on the Mathematics of Biodiversity

CRM, Bellaterra

July 2 to 6, 2012

Scientific Committee

Ben Allen, Harvard University
Silvia Cuadrado, Universitat Autònoma de Barcelona
Tom Leinster, University of Glasgow
Richard Reeve, University of Glasgow
John Woolliams, University of Edimburgh

Speakers

Ben Allen, Harvard University
John Baez, University of California, Riverside
Neil Brummitt, Natural History Museum, London
Anne Chao, National Tsing Hua University
Christina Cobbold, University of Glasgow
Elisabeth Gillet, Universität Göttingen
Hans-Rolf Gregorius, Georg-August-Universität Göttingen
Lou Jost, Independent researcher
Tom Leinster, University of Glasgow
Alison Mather, The Wellcome Trust Sanger Institute
Louise Matthews, University of Glasgow
Hans Metz, Universiteit Leiden
Sandrine Pavoine, Muséum National d'Histoire Naturelle
Richard Reeve, University of Glasgow
Carlo Ricotta, Sapienza Università di Roma
William Sherwin, University of New South Wales
Mike Stear, University of Glasgow
Simon Willerton, University of Sheffield
John Woolliams, University of Edinburgh

Acknowledgements: This research program was supported in part by the UK's Biotechnology and Biological Sciences Research Council (BB/J020567/1) and Ministerio de Ciencia e Innovación (MTM2011-13211).

Contents

1. Practical Information

Lecture room: The workshop will take place in the "Auditori" hall of the CRM. The lecture room is equipped with blackboard and projector.

Secretariat: The Secretariat of the CRM will be available to the participants Monday through Friday from 9:00 am to 2:30 pm. Please, note that the Administration will not be available in the afternoons.

Computer facilities: The computer space of the CRM will be available for the participants of the Conference.

CRM wireless information:
Wireless network password: crmwifikey
Security: WPA-only
Cipher Type: TKip
Authentification code: PSK

CRM computers information:
Username:crmactivities
Password: Crm2012

Additional meeting rooms can be reserved upon request.

Social events: We have organized a guided visit and dinner for Wednesday, July 4th, in the afternoon. Registration to Sagrada Família basilica was offered in advance and the list is already closed. However, if you didn't register, and wish to participate, please talk to Neus Portet to see if there are available entrance tickets. Registration to dinner will close on July 3rd at 1 pm. Talk to Neus to sign in for dinner.

Lunch: There are a few restaurants on campus. You can check the information at `http://servei-restauracio.uab.cat/`

Picture: A group picture will be taken on Thursday, July 5 before the coffee break. We will inform you of the place to meet. The picture will be posted on the workshop's web page.

Questionnaire: Following the directions of the CRM Governing Board, we give a questionnaire to all the people participating in activities at the CRM in order to assess their level of satisfaction. The questionnaire is anonymous and not mandatory, but we would greatly appreciate it if you could answer the questions and return it to us. Thank you for your cooperation.

Local emergency numbers: General emergency (police, ambulance, fire-fighters) call 112.

Safety in Barcelona: Although Barcelona is a safe city, please be aware that there is a problem with pickpockets, especially around tourist areas: La Rambla, Plaça Catalunya, Barcelona Airport, major metro and train stations, famous buildings, etc. Be sure to keep your belongings with you at all times, be alert, and be wary of unusual situations.

Further information

Centre de Recerca Matemàtica
Ms. Neus Portet
`http://www.crm.cat`
nportet@crm.cat
Tel.: 93 586 40 86

2. Schedule of the Conference

Monday, July 2	
09:00 – 09:30	Registration and welcome
09:30 – 10:30	Lou Jost, Independent researcher *Why biologists should care about the mathematics of biodiversity*
10:30 – 11:00	Coffee break
11:00 – 12:00	Special session: Who are we? What have we been doing?
12:15 – 13:15	Anne Chao, National Tsing Hua University *Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers*
13:15 – 15:00	Lunch break
15:00 – 16:00	Carlo Ricotta, University of Rome "La Sapienza" *On the equivalent number of partially distinct species: Some theory and a practical example*
16:00 – 16:45	Contributed talk: Andrés Baselga, Universidad de Santiago de Compostela *Partitioning beta diversity into turnover and nestedness components*
17:00 – 17:45	Contributed talk: Yoni Gavish, Ben-Gurion University of Negev *Determinants of true alpha, beta, and gamma diversities of spiders in a fragmented landscape - a combinatorial network approach*

Tuesday, July 3	
09:30 – 10:30	Hans Metz, Universiteit Leiden *Some geometrical principles underlying longer term evolution*
10:30 – 11:00	Coffee break
11:00 – 12:00	Alison Mather, Wellcome Trust Sanger Institute *The diversity of antimicrobial resistance – a different perspective on comparing microbial populations*
12:15 – 13:15	Christina Cobbold, University of Glasgow *Measuring biodiversity: the importance of species similarity*
13:15 – 15:00	Lunch break
15:00 – 16:00	Tom Leinster, University of Glasgow *Maximizing diversity: you can please all of the people all of the time*
16:00 – 16:45	Special session: Who are we? What have we been doing?
17:00 – 17:45	Contributed talk: Núria Teixidó, Universitat de Barcelona *Biodiversity patterns in marine rocky communities of the Mediterranean Sea; some insights from geographic and temporal network approach*

Wednesday, July 4	
09:30 – 10:30	William Bruce Sherwin, University of New South-Wales *Entropy and information approaches to genetic diversity and its expression: Genomic geography*
10:30 – 11:00	Coffee break
11:00 – 12:00	John Woolliams, University of Edinburgh *Diversity in livestock: from entropy to squares*
12:15 – 13:15	Neil Brummitt, Natural History Museum, London *The sampled red list index for plants: a baseline biodiversity assessment*
13:15 – 15:00	Lunch break
15:00 – 22:00	Group activity (Times are approximate. Further information to be given in separate document)

Thursday, July 5	
09:30 – 10:30	John Baez, University of California, Riverside *Diversity, information geometry and learning*
10:30 – 11:00	Coffee break
11:00 – 12:00	Hans-Rolf Gregorius, Georg-August Universität Göttingen *Classifying measures of biological variation*
12:15 – 13:15	Elizabeth Gillet, Universität Göttingen *Measuring effects of gene associations on differentiation*
13:15 – 15:00	Lunch break
15:00 – 16:00	Louise Matthews, University of Glasgow *Measuring, maintaining and maximising mhc diversity*
16:00 – 16:45	Mike Stear, University of Glasgow *The major histocompatibility complex: Quantifying, explaining and exploiting the most diverse region of the mammalian genome*
17:00 – 17:45	Simon Willerton, University of Sheffield *Magnitude and other measures of metric spaces*

Friday, July 6	
09:30 – 10:30	Sandrine Pavoine, Muséum national d'Histoire naturelle, Paris *Partitioning and visualizing biodiversity: when species dissimilarity matters*
10:30 – 11:00	Coffee break
11:00 – 12:00	Richard Reeve, University of Glasgow *Quantifying antigenic diversity - the nonconcept of serotype*
12:15 – 13:15	Ben Allen, Harvard University *Phylogenetic entropy: Incorporating community phylogeny into the Shannon index*
13:15 – 15:00	Lunch break
15:00 – 16:00	Lou Jost, Independent researcher *Challenges and open questions in diversity theory*
16:00 – 17:45	Discussion

3. Abstracts of the Speakers

Ben Allen

Phylogenetic entropy: Incorporating community phylogeny into the Shannon index.

Abstract: I will present phylogenetic entropy, an index I developed in 2009 to characterize the biodiversity of an ecological community. Phylogenetic entropy generalizes the Shannon index to incorporate the phylogenetic tree structure of the community. In general, this index increases with the number of species, the evenness of their distribution, and the degree of phylogenetic divergence among species. Phylogenetic entropy satisfies a number of desirable mathematical properties, including a hierarchical decomposition property and "weak species monotonicity"—meaning that it always decreases if a rare species is eliminated. I will illustrate the application of this index to phyllostomid bat communities. I will also discuss how phylogenetic entropy fits within the larger project of establishing diversity indices that reflect species' relatedness as well as their distribution.

Contact address: benjcallen@gmail.com

John Baez

Diversity, information geometry and learning.

Abstract: As is well known, some measures of biodiversity are formally identical to measures of information developed by Shannon and others. Further, the replicator equation in evolutionary game theory is formally identical to a process of Bayesian inference studied in the field of machine learning. Thus, in this simple model, a population of organisms can be thought of as a hypothesis' about how to survive, and natural selection acts to update this hypothesis according to Bayes' rule. This idea has been studied by Marc Harper and elaborated using ideas of information geometry. The question thus arises to what extent natural changes in biodiversity can be usefully seen as analogous to a form of learning. However, some of the same mathematical structures arise in the study of chemical reaction networks, where the increase of entropy, or more precisely decrease of free energy, is not usually considered a form of 'learning'. We report on some preliminary work on these issues.

Contact address: baez@math.ucr.edu

Neil Brummitt

The sampled red list index for plants: a baseline biodiversity assessment.

Abstract: The scale of the global biodiversity crisis means that international efforts to identify, conserve and monitor threatened species must be carried out with greater urgency than ever. This is especially true for plants, with only 4% of species worldwide previously assessed and on the IUCN Red List. The Sampled Red List Index (SRLI) for Plants is an attempt to address this deficiency for a small, randomly-selected sample of plant species from around the world. In the absence of detailed information on the status and trends of populations of threatened species, herbarium specimens are often the most easily accessible data on which to base an assessment of conservation status. This is the approach adopted for this work, and from this baseline further studies have applied the EDGE (Evolutionarily Distinct and Globally Endangered approach to plant groups assessed for the SRLI project, and assessed patterns of functional and phylogenetic diversity for the sample of monocot species. This talk will present the results from these different approaches, and discuss the international biodiversity policy context in which the work has been carried out.

Contact address: `n.brummitt@nhm.ac.uk`

Anne Chao

Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers.

Abstract: Hill numbers or "effective number of species" resolve many of the interpretational problems caused by classical diversity indices (e.g. Shannon entropy and the Gini-Simpson index), but they treat all species as equally distinct. Chao, Chiu and Jost (2010) generalized Hill numbers to take into account phylogenetic or functional differences between species. Here we show how to multiplicatively partition the new phylogenetic gamma diversity into independent alpha and beta components. The resulting beta diversity (ratio of gamma to alpha) measures the heterogeneity of the region, in units of effective number of completely distinct assemblages. The difference between gamma and alpha (lineage excess) quantifies the absolute phylogenetic differentiation in units of amount of lineage lost between a regional and a local scale. The phylogenetic beta component measures pure differentiation among assemblages and thus can be used to construct relative similarity or differentiation measures. Both phylogenetic beta diversity and lineage excess lead to the same family of relative similarity or differentiation measures in the range $[0, 1]$. The resulting family of similarity measure is a phylogenetic generalization of the Sørensen, Jaccard, Horn, and Morisita-Horn similarity measures. We also compare our approach with traditional phylogenetic additive decomposition based on quadratic entropy

(an extension of the Gini-Simpson index) or phylogenetic entropy (an extension of Shannon entropy). Hypothetical and real examples are used for illustration. (This is a joint work with C. H. Chiu and Lou Jost.)

Contact address: chao@stat.nthu.edu.tw

Christina Cobbold
Measuring biodiversity: the importance of species similarity.

Abstract: Biodiversity is associated to properties of ecosystem function and productivity or at the molecular level it can describe genetic diversity with connections to quantities such as host-pathogen fitness. Diversity is clearly an important concept and it begs the question how should one measure diversity. There are literally dozens of measures of diversity in the literature, perhaps one of the most common examples being species richness or simply the number of species in the community concerned. Even such a simple definition raises questions, what defines a species? Sometimes this question is simple to answer, but for microbial communities, for example, there is no simple definition of a species and yet we still wish to quantify diversity.

To address this issue we present a natural family of diversity measures which not only takes into account relative abundance of species, but also accounts for differences or similarities between species. This latter point allows us to deal with communities where the notion of species is unclear.

We demonstrate that our new measure of diversity is not simply an addition to the already long list of indices: instead, a single formula subsumes many of the most popular indices, including Shannon's, Simpson's, species richness, and Rao's quadratic entropy. These popular indices can then be used and understood in a unified way, and the relationships between them are made plain. The new measures are, moreover, effective numbers, so that percentage changes and ratio comparisons of diversity value are meaningful.

We illustrate these measures with the use of diversity profiles, a simple graphical tool which provides information about the shape of a community and how its diversity changes according to the importance that is placed on rare or common species. The comparison is simple, but reveals important insights into a community's diversity while also being transparent about how diversity is defined, so rather than simply settling for one measure of diversity many can be considered and thus more information about the community can be revealed.

Contact address: Christina.Cobbold@glasgow.ac.uk

Elisabeth Gillet
Measuring effects of gene associations on differentiation.

Abstract: Genetic differentiation quantifies the dissimilarity among the populations that make up a metapopulation for their distributions of the states of a

genetic trait. It is bounded, assuming its maximum only when all populations are completely dissimilar. Since populations may differ not only in their allele distributions but also in their gene associations, it may be misleading to assess differentiation using gene-pool frequencies alone. Selection, in particular, may produce dissimilarities at the single- or multilocus level of genetic integration, since selection acts on phenotypes, the coding of which often involves interactions among the genes at multiple loci. Homologous interactions may be observable as deviations from Hardy-Weinberg proportions among individual loci, while non-homologous interactions appear as gametic disequilibrium among loci. Such deviations from random gene association may also result from drift in small populations (bottleneck). Appropriate measures of differentiation must therefore be applicable to all levels of genetic integration by considering not only the frequencies of the genetic types but also the genic difference between types. For two populations, the measure $\triangle$ of genetic distance quantifies the minimum amount of change necessary in one population to make its frequency distribution match that of the other. It could be proven that $\triangle$ does not decrease when the integration level is increased from the gene-pool over the mean single-locus to the multilocus level. $\triangle$ remains equal at the single-locus level if both populations show equal inbreeding coefficients, and at the multilocus level if the single-locus genotypes are in gametic equilibrium. An increase in $\triangle$ would therefore indicate that the two populations differ in their forms of gene association. Generalization of $\triangle$ to a measure of differentiation in metapopulations comprising higher numbers of populations can be done in different ways. A dispersive measure of differentiation calculates a mean $\triangle$ over all pairs of populations, while the complementary differentiation measure $\triangle_{SD}$ reflects the symmetric set difference between each population and its complement. The effects of gene associations on these measures are examined.

Contact address: `egillet@gwdg.de`

Hans-Rolf Gregorius

Classifying measures of biological variation.

Abstract: Terms such as “diversity” and “differentiation” are applied to a variety of features of biological variation, with the result that the original meaning of the terms has become somewhat blurred. In some cases, it has been demonstrated repeatedly that widely used indices of “differentiation” are incompatible with their claimed meaning. In other cases, the terms are stretched to a degree that obscures their original determination. Concepts such as “diversity between communities” or beta-“diversity” are examples. To ameliorate this, more recent attempts to include variable differences into the assessment of variation propose “effective numbers” as suitable descriptors of diversity.

Since the observation of variation implies the observation of differences, it seems reasonable and timely to look for possibilities to distinguish and classify measures

of variation on the basis of a general notion of difference between objects. Having defined the latter, the notions of dispersion (spread), diversity, differentiation (between communities), and partitioning of variation will be addressed in contexts of difference. Partitioning of variation will be considered with respect to its two basic appearances: division of variation among communities, and partitioning of total diversity into components within and between communities. Opportunities to apply the concept of effective number to these notions will be outlined. The dual perspective of differentiation, where communities may differ for the trait states of their members and trait states may differ for the community membership of their carriers, will be pointed out to help in resolving some of the common controversies over the term.

Contact address: gregorius@gwdg.de

Lou Jost
Why biologists should care about the mathematics of biodiversity.

Abstract: For many years ecologists, conservation biologists, and geneticists have used diversity measures uncritically, without checking to see whether these measures really work as expected. They have attempted to construct measures of compositional similarity and differentiation from their diversity measures, without understanding the mathematical connection between diversity and compositional similarity. Many of these measures do not support the interpretations usually applied to them. As a result, the biological literature is littered with mistaken analyses and invalid reasoning, not merely in empirical studies but even in foundational theoretical work. This is especially serious in population genetics, where the most important measure of genetic divergence Gst or Fst) is based on a pair of mathematical misconceptions. I identify some of the mathematical properties implicit in biological reasoning about diversity and similarity, show how some of the most commonly used measures lack these properties and lead to contradictions, and give measures which really have the properties biologists require. The new tools often give very different answers than the old tools. I apply them to a fundamental biological process, the evolution of genetic divergence in subdivided populations. I show that the current understanding of this process is mistaken, because geneticists didn't pay enough attention to the mathematics of diversity. I give the correct version, confirmed by simulations and empirical tests.

Contact address: loujost@yahoo.com

Lou Jost
Challenges and open questions in diversity theory.

Abstract: Many important questions remain unsolved in diversity theory. These include some issues that will have arisen during the last few weeks of discussions

here at the CRM, and issues that have appeared previously in the literature. I will focus on issues of practical interest to biologists.

Contact address: loujost@yahoo.com

Tom Leinster
Maximizing diversity: you can please all of the people all of the time.

Abstract: Different ecologists, shown two communities, will make different judgements on which is the more diverse. One axis of difference is the relative importance attached to rare and common species: while one person might prioritize conservation of rare species, another might prioritize overall community balance. This spectrum of opinions is captured by the diversity measures known as the Hill numbers, which involve a parameter q specifying one's position on the spectrum. However, the Hill numbers are based on a crude model in which the varying similarities between species are ignored. It is as if distinct species have nothing whatsoever in common. Christina Cobbold's talk will show how to repair this defect, extending the Hill numbers to take inter-species similarity into account. I will then show how to maximize diversity (in this theoretical sense). In other words, given a list of species of known similarities, I will describe the frequency distribution with the highest possible diversity. The big surprise is this: there is a single frequency distribution that maximizes diversity from all points of view simultaneously. No matter whether one's priority is rare species (low q) or common species (high q), this distribution is optimal. Moreover, the value of the maximum diversity is the same for all q: so any list of species of known similarities has an unambiguous maximum diversity.

Contact address: Tom.Leinster@glasgow.ac.uk

Alison Mather
The diversity of antimicrobial resistance – a different perspective on comparing microbial populations.

Abstract: Bacterial infections that are resistant to antimicrobial drugs threaten the health of humans and animals worldwide. Historically, investigation into this issue has typically focused on examining the prevalence of resistance to individual antimicrobials. However, this does not take into account the potential co-dependence of resistance, as multiple resistance genes can be located on the same genetic element, such as a plasmid. Our work takes a different perspective, considering the full spectrum of resistance exhibited by a bacterium. We compare antimicrobial resistance (AMR) in different populations by examining the diversity of resistance patterns, or profiles, in each population. We do this by applying the concepts of diversity metrics, developed to compare species, to the resistance profile data, examining diversity measures that include all weightings

of the importance of the number of unique profiles and of the abundance of each profile to each measure. We compare AMR diversity both at the resolution of the observed sample of bacterial isolates, and, utilising the existence of a maximum number of resistance profiles to a given number of tested antimicrobials, at the resolution of the underlying and unobserved populations of bacteria. We have applied these methods to the comparison of AMR of bacteria from co-located animals and humans, the comparison of different surveillance systems for the detection of emerging or rare resistances, and to investigate the associations of host age and antimicrobial exposure with AMR diversity within dairy cattle. These approaches provide a different perspective on AMR, one that can contribute to a greater understanding of the ecology of resistance. However, methodological challenges remain, including the determination of appropriate sample sizes for robust comparison using these metrics between two or more populations.

Contact address: am24@sanger.ac.uk

Louise Matthews
Measuring, maintaining and maximising mhc diversity.

Abstract: (Joint work with Richard Reeve, Mike Stear, and Thorsten Stefan.) A key element of the immune response in mammals is encoded by a region of the genome called the major histocompatibility complex (mhc). The mhc is highly polymorphic, displaying far greater numbers of gene variants (alleles) than any other region of the genome. Why such diversity exists, however, is a source of continued debate. A commonly promoted explanation hinges on the observation that individuals with two different alleles (heterozygotes) can recognise more pathogen antigens than individuals with two identical alleles (homozygotes). Thus, populations with more alleles and therefore more heterozygotes are assumed to be fitter than those with few alleles. Whether this mechanism is sufficient to explain the degree of diversity we observe in the natural world is an area of active investigation. We believe that current mathematical models of the maintenance of mhc diversity give misleading predictions of population fitness and allele abundance because they do not account for the distance between alleles in terms of the repertoire of pathogen antigens recognised. Functional diversity offers a resolution to this debate. We propose new models for the maintenance of mhc diversity that capture both allele fitness and allele distance. By capturing diversity in a way that is biologically meaningful, these models provide i) more realistic predictions of the evolution and persistence of allele diversity, ii) measurements of mhc diversity which can be used to compare population fitness, and iii) a metric against which to assess the impact of selective breeding or conservation schemes.

Contact address: Louise.Matthews@glasgow.ac.uk

Hans Metz

Some geometrical principles underlying longer term evolution.

Abstract: My last direct expertise with biodiversity problems was 40 year ago; since then I have only observed how the term is used in my scientific environment. Therefore, I surmise that I was asked to participate for my expertise in concept engineering and evolutionary population biology. Apparently recently some pleasing fundamental progress has been made with regard to the consistency and interrelations of diversity measures. The former is a necessary precondition for them to be be useful. However, it is less clear to me for which tasks the measures have been designed. Of course, they can be used in a descriptive manner, and to shore up conservation policies, given that politicians and lawgivers have already been sold on the idea that biodiversity is worthwhile. However, my personal interest is mostly in another side of the scientific process, to wit, in how measures can be used for phrasing empirical or theoretical rules. Although diversity is a bit of a buzzword in the literature on community dynamics and evolution, I so far did not find any analytical paper where the term was used more sophisticatedly than as referring to numbers of species (where I mean with analytical that the paper tried to uncover a connection between diversity and some other phenomenon). Moreover, there appears to be little consistency in the discerned relationships (e.g. Ives (2007) Science 317: 58-62). Apparently, there then is rather a disconnect between the sophistication of the representational techniques and the relative lack of their use as hard tools. This is not meant as a criticism (for this I too much enjoyed reading up on the diversity math!), but rather as a pointer to potential opportunities. Recently there has been spate of attempts at evolving communities in silico in order to get a handle on the difference between evolved and artificially assembled community models. Although these attempts clearly are seductive, this work so far leaves the feeling that we still miss good tools for recovering potential generalities from the model outcomes. Many of these models stand on their own, although their originators often try to justify them with a reference to some poorly justified universality conjecture. However, some models are based more on established biological mechanisms. These fit with my own interest in uncovering general mathematical principles for the development and loss of trait diversity on meso- and macro-evolutionary scales from population dynamical and evo-devo considerations. In the remainder of the talk I will summarize some of those principles in the hope that this may help you connect a little to this context.

Contact address: `j.a.j.metz@biology.leidenuniv.nl`

Sandrine Pavoine

Partitioning and visualizing biodiversity: when species dissimilarity matters.

Abstract: The concept of diversity encompasses all kinds of variability and is strongly dependent on the concept of being dissimilar. It has been studied in many disciplines, each having its own vocabulary and rules. This talk adopts an ecological viewpoint but it recalls that despite differences in vocabulary and rules, analyses of diversity can be unified by similarities in the structure of the analysed object (be they language words in a text, cultural expressions in a country, alleles in a genome, species in a community, habitats in a landscape). In ecology, biodiversity has been usually measured by counting species and comparing their abundances. Information on species number and abundance are now complemented by how dissimilar species are. Dissimilarities among species are defined using their biological traits, even acoustic traits, and phylogenetic/taxonomic positions. The addition of species? characteristics aims to provide a better understanding on the ecological and evolutionary mechanisms that determine which species co-exist in an area. The first part of the talk presents different ways of partitioning biodiversity in ecology depending on whether space, ecological time or evolutionary time is of interest. It suggests direction for future research were critical methodological steps have not yet been solved. The second part highlights connections between diversity partitioning approaches and factorial analyses. Factorial analyses display simple and multi-table data sets in an appropriate Euclidean space with the ultimate objective of summarizing data into spaces of few dimensions that can be more easily interpreted. Both parts are illustrated with the analysis of real data sets.

Contact address: `pavoine@mnhn.fr`

Richard Reeve

Quantifying antigenic diversity – the nonconcept of serotype.

Abstract: Boyd Orr Centre for Population and Ecosystem Health Institute of Biodiversity, Animal Health and Comparative Medicine College of Medical Veterinary and Life Sciences University of Glasgow UK Viruses and bacteria are classified antigenically into serotypes (or serovars) according to variations in their surface antigens. These changes are detected serologically, with pathogens usually placed in distinct serotypes when serum raised against one pathogen does not cross-react with another. This classification system does not handle continuous small changes in antigenicity well, allowing pathogens that do not cross-react to remain in the same serotype when they are connected by a chain of cross-reacting pathogens. As a result, a new pathogen can only be correctly classified into an existing or new serotype if we have prior knowledge of all other related strains. This makes the system vulnerable to missing information on intermediate pathogens,

which may cause misidentification of a new serotype, or even to the addition of rare and epidemiologically irrelevant intermediate isolates that make practically distinct serotypes appear to join. An additional weakness is that the system does not take into account the antigenic distance between pathogens, which determines whether infection with or vaccination against one will protect against subsequent infection by another. It therefore ignores the important practical distinction between narrow serotypes comprising antigenically similar pathogens and broad, antigenically diverse, serotypes. We are investigating the use of serotype diversity measures, incorporating serological cross-reactivity as a measure of distance between pathogens, to count the "effective number of serotypes"- defined either as the number of vaccines required to cover one or a group of serotypes by vaccination, or as the number of distinct strains, infection with which will not protect against the others. We believe this number will be both more stable and robust to missing data and of more practical value than the current crude labelling scheme.

Contact address: `Richard.Reeve@glasgow.ac.uk`

Carlo Ricotta
On the equivalent number of partially distinct species: Some theory and a practical example.

Abstract: Many applications of diversity indices are only appropriate if they are first converted into their equivalent number of species. To maintain the connection to the original definition of an equivalent number of species for traditional diversity measures, Ricotta & Szeidl defined the species equivalent of Rao's quadratic entropy Q as the number of equally likely and maximally distinct species needed to produce the given value of Q. Here, I generalize the notion of an equivalent number of species for the Rao diversity to partially distinct species. The biological meaning of this proposal is illustrated with one dedicated case study in sand dune communities in Italy.

Contact address: `carlo.ricotta@uniroma1.it`

William Sherwin
Entropy and information approaches to genetic diversity and its expression: Genomic geography.

Abstract: This talk highlights advantages of entropy-based genetic diversity measures, at levels from gene expression to landscapes. Shannon's entropy-based diversity is the standard for ecological communities. The exponentials of Shannon's and the related "mutual information" excel in their ability to express diversity intuitively, and provide a generalised method of considering microscopic behaviour to make macroscopic predictions, under given conditions. The hierarchical nature of entropy and information allows integrated modeling of diversity

along one DNA sequence, and between different sequences within and among populations, habitats, species, etc. The ultimate aim is to identify the formal connections between genetic diversity and the flow of information to and from the environment. We have developed predictive equations for Shannon summaries of neutral genetic variation under various population histories and mutation modes. These equations are analogous to the familiar predictive equations for heterozygosity, which were then applied to neutral ecological communities by Hubbell. Moreover, it is important to note that when considered on a base-by-base scale, entropy-based measures of diversity also allow use of other mathematical methods not devised specifically for entropy and information. This further extends the flexibility and utility of entropy-based approaches, especially when trying to develop diversity measures that are sensitive to species similarity.

Contact address: w.sherwin@unsw.edu.au

Mike Stear

The Major Histocompatibility Complex: Quantifying, explaining and exploiting the most diverse region of the mammalian genome.

Abstract: The Major Histocompatibility Complex is a genetic region whose genes play a key role in the immune response against pathogens and parasites. It is also the most diverse region of the mammalian genome and specific genetic variants confer increased resistance or susceptibility to a variety of different diseases. The high levels of diversity are probably a consequence of natural selection for disease resistance. However, despite 3 sets of Nobel prizes and over 20,000 papers there is no clear consensus on how to compare diversity across genes, chromosomes, individuals, subpopulations or species. There is no consensus on the mechanisms that maintain diversity or even how this diversity could be managed to improve the disease resistance of livestock. Here we will present our first steps in quantifying genetic diversity, explaining how selection maintains high levels of diversity and breeding livestock with improved resistance to disease.

Contact address: Michael.Stear@glasgow.ac.uk

Simon Willerton

Magnitude and other measures of metric spaces.

Abstract: Leinster introduced the notion of the magnitude of a metric space, which to a set of points with distances associates some notion of size. He and Cobbold also introduced a family of diversity measures which depend on both distance between species and relative abundance of species.

In many situations the magnitude of a set of species tells you the maximum value these diversity measures can reach as the relative abundances of the species vary. In some situations a variant of the magnitude is required, however. In this talk I will show how Leinster and Cobbold's diversity measures can be used

to construct other notions of size for metric spaces and compare them with the variants of magnitude mentioned above. This will be illustrated with various examples.

Contact address: S.Willerton@sheffield.ac.uk

John Woolliams
Diversity in livestock: from entropy to squares.

Abstract: Livestock breeding has developed a mix of tools to manage diversity in its species. Concerns over diversity in livestock are primarily concerned with genetic rather than spatial diversity, although the linking of breed to a region is both a source of support for the maintenance of breed diversity, and a risk of loss of diversity through disease or other localised disasters. Diversity in livestock is considered within a species rather than between species, however this variation has a structure formed by breeds. The breeds are positively assortatively mated, so that variation can be partitioned between breeds and within breeds. The resource for biodiversity is therefore variation between breeds and within breeds, although there is gene flow between breeds and also structured crossing to make production systems more efficient.

The criterion for management is the rate of change of this diversity. Of particular interest is the management within breeds which is based upon the rate of loss of genetic variation, which is in turn is measured by the rate of inbreeding, $\triangle F$. In population genetics, the term effective population size (N_e) is more often used, but the only meaningful definition of N_e is as the transform $(2\triangle F)^{-1}$. There is a fundamental theorem in quantitative genetics which shows $\triangle F = 1/4 \sum r_i^2$, where r_i is the long term genetic contribution of individual i to the future gene pool and the sum is over all ancestors in a generation. Thus management of diversity in populations depends on the sum of squared proportions not on the entropy. Many fundamental properties of population besides rate of the loss of variance are determined by Ne (i.e. $\triangle F$) such as magnitude of changes in allele frequencies, speed of differentiation, the maintenance of linkage associations etc. In breeding programmes $\triangle F$ can be regarded as a measure of risk in genetics. Therefore a natural extension of this to the measurement of diversity across a region is to replace the use of entropy as in $r_i \ \log r_i$ by the use of squares r_i^2. Such a step allows a natural decomposition of total diversity in a region, into sub-regional components, into localities, species within localities and diversity within species. The relationships between localities can also be assessed by crossproducts. These steps are analogous to decompositions into means variances and covariances.

Contact address: john.woolliams@roslin.ed.ac.uk

4. Abstracts of the Contributed Talks

Andrés Baselga

Partitioning beta diversity into turnover and nestedness components.

Abstract: The strict sense definition of beta diversity is the ratio between gamma and alpha diversities, i.e. the factor to which the diversity of a region exceeds the mean diversity of local sites within the region. Gamma diversity can be different to alpha diversity if, and only if, local sites differ in species composition. Therefore, the actual parameter determining beta diversity is the degree to which species composition changes from site to site.

Although the intuitive concept of "change in species composition" is apparently straightforward, two different phenomena can produce differences in species composition between two sites. The first phenomenon is the replacement of some species by others from site to site, a concept that has been termed spatial turnover. The second phenomenon is nestedness, a pattern characterised by the poorest site being a strict subset of the richest site. In this case both sites have obviously different species composition (i.e. the richest site has unique species not present in the poorest site), but no species is replaced by other. Therefore the total dissimilarity between two assemblages (i.e. a monotonic transformation of beta diversity) can be partitioned into two components: dissimilarity due to species replacement and dissimilarity due to nestedness.

This framework has been developed for dissimilarities derived from beta diversity of order $q = 0$ (diversity measured as species counts; Hill numbers for $q = 0$), and it is extended here to dissimilarities based on absolute abundances (i.e. Bray-Curtis dissimilarity). Finally, the applicability of this partitioning framework to beta diversity of order $q > 0$ is explored, and future lines of research are suggested.

Contact address: `andres.baselga@usc.es`

Yoni Gavish

Determinants of true alpha, beta, and gamma diversities of spiders in a fragmented landscape - a combinatorial network approach.

Abstract: Understanding the ecological and environmental determinants of biodiversity patterns in patchy and fragmented landscapes is of the greatest challenges of ecology. However, methodological constrains on the sampling effort usually limit the number of landscapes that can be practically sampled. Consequently, sample size at the landscape-scale is low, making the inference at the landscape scale problematic because the analysis mainly reflects the patch-scale while similar analysis at the landscape is rarely doable. One possible approach to overcome the lack of repetition at the landscap-scale is using the same landscape as pseudo-repetition of itself. For example, if we sample n patches within

a landscape, we can list all possible combinations of $n-1$ patches. Each such combination is considered as one combinatorial network. Next, we can calculate various response (e.g., beta diversity) and explanatory (e.g., minimum spanning tree) variables for each combinatorial network. Finally, we can explore the covariance between the explanatory and the response variables. We can then repeat the entire procedure for all combinations of $1, 2, \ldots, n$ patches.

We employed the combinatorial network approach for spiders in the fragmented landscape of Southern Judea Lowland, Israel. We have sampled spiders in 12 patches and one relatively large continuous area within a 3.2 km^2 landscape. Each sample was 0.5×0.5 m^2 and contained one of 9 apriori defined microhabitats. Spiders and other arthropods were sampled using the vacuum option of a leaf blower with a mesh (0.5 mm) sleeve inserted within the suction tube. All spiders larger than 0.5 mm in total length were identified to the species or morphospecies level. All other arthropods (potential prey) were identified to the order level. A total of 218 samples were taken, resulting with 4047 spider individuals (147 species) and 9779 other arthropods. The relative cover of the nine microhabitats in each patch was quantified using 110 line transects, 20 m long each.

Our main ecological aim was to explore the relative importance of explanatory variables from five main themes (sampling, heterogeneity, area, connectivity and potential prey diversity) on alpha, beta and gamma diversities of spiders. More specifically, we explore how the relative importance of these five main themes in explaining spiders alpha, beta and gamma diversities change with the number of patches in the network and with two spatial types of the network – archipelagic and mainland-island networks. We first listed all possible combinatorial networks of $1, 2, \ldots, 13$ patches. We then subdivided all networks to those that contain the large unfragmented area (mainland-island systems) and those that did not (archipelagic systems). We further subdivided all networks from a given spatial network type according to the number of patches in the network. For each possible combinational network we calculated three response variables – the true alpha, beta and gamma diversities of spiders – using the numerical equivalent of Shannon entropy. We used the total number of individuals sampled within each patch as weights (i.e., unequal weights).

For the sampling theme, we estimated four explanatory variables, the first being the total number of samples in the network. Next, we calculated the true alpha diversity of the sampling, using the proportion of samples of each microhabitat in each patch for the Shannon entropy. We used the proportion of samples in each patch from the total number of samples in the network as weights. The true alpha diversity of sampling gives the effective number of microhabitats that were actually sampled in each patch. The true beta diversity of the sampling reflects the effective number of patches with uniquely sampled microhabitats. The true gamma diversity of sampling provides the effective number of microhabitats that were sampled within the network. For the heterogeneity theme, we used the

line transect data to estimate the true alpha, beta and gamma diversity of heterogeneity. Here, Alpha diversity reflects the effective number of microhabitats within each patch, beta diversity reflects the effective number of patches with completely unique microhabitats and gamma diversity reflects the effective number of microhabitats at the entire network. For the area theme, we calculated the total and average area of the patches, as well as the true alpha diversity of the area, which serves here as a measure of evenness (i.e., values close to 1 indicate the network is dominated by one large patch). For the connectivity theme, we quantified the length of the network's minimum spanning tree. In addition, we calculated the area of the minimum convex hull, and used the ratio between the total area of patches and the area of the convex hull as a measure of network's spatial clumping. Similar to the procedure for spiders, we estimated the true alpha, beta and gamma diversities of the potential prey.

We repeated the following statistical procedure for each response variable for each subset of the entire set of combinatorial networks (number of patches and spatial network type), using an information theory framework. We first listed all possible linear models in all possible combinations of the explanatory variables (without interactions). We then used GLM (normal error distribution with identity link) for each model against each of the three response variables. We recorded the AICc weight of each model. Next, for each explanatory variable we calculated the relative importance weights by summing the AICc weights of all models that contained the focal explanatory variable. Next we used linear regression and permanova for each explanatory variable to explore the change in its importance with the number of patches and the spatial type of the network. Current results suggest that using Jost partitioning and true diversities may highly contribute to our ecological understanding of complex landscapes. For example, spider beta diversity (but not alpha diversity) increases with the length of the minimum spanning tree, which is the network equivalent of distance decay of similarity. In addition, alpha diversity of spiders is strongly correlated with alpha diversity of potential prey, while spiders' beta diversity is mostly correlated with beta diversity of prey.

In summary, this work emphasizes the ecological application of Jost partitioning and true diversities not only for the response variable, but also for explanatory variables. We show one possible usage of true diversities within a combinatorial network approach, which may allow inference at the landscapescale, under methodological sampling constrains. In addition, our analysis may shed light on relatively unexplored biodiversity patterns such as the effect of the number of patches and their spatial configuration on alpha, beta and gamma diversities in mainland-island and archipelagic like systems.

Contact address: gavishyoni@gmail.com

Núria Teixidó

Biodiversity patterns in marine rocky communities of the Mediterranean Sea; some insights from geographic and temporal network approach.

Abstract: The Mediterranean Sea is considered a marine biodiversity hotspot, harboring approximately 10% of world's marine species while occupying only 0.82% of the ocean surface. Unfortunately, the impacts of human activities are proportionally stronger in the Mediterranean than in the other seas, raising concerns regarding threats to the conservation of the rich Mediterranean biodiversity. Coralligenous outcrops, which are hard bottoms of biogenic origin that thrive under dim light conditions, are among the habitats faced with major threats in the Mediterranean Sea. These outcrops are highly diverse (harboring approximately 20% of Mediterranean species) and exhibit great structural complexity. The species that characterize coralligenous seascapes are encrusting calcareous algae, sponges, cnidarians, bryozoans and tunicates. Some of the engineering species in these environments are long-lived, and their low dynamics make coralligenous outcrops exceptionally vulnerable when faced with sources of strong disturbances, such as destructive fishing practices, pollution, invasive species or mass mortality outbreaks. Despite the ecological, aesthetic and economic value, coralligenous outcrop biodiversity patterns are poorly understood both at regional and temporal scales.

We present preliminary data from an ongoing study to estimate the diversity of macro-species of algae and invertebrates (mainly sponges, cnidarians, bryozoans and tunicates) through photographic surveys across large spatial ($36^0 - 43^0$ latitude, > 1000 Km) and temporal scales (5 years) of coralligenous communities. We use a hierarchical sampling procedure to measure species richness among regions, regions within localities, and localities within sites as a standardized monitoring network. At each site, 3 replicates of 8 photographic quadrats ($25 * 25$ cm) are taken according to the minimum sampling area for these communities. For each site, around 50 benthic species are recorded. The large spatial approach of this study is optimal to analyse the variability over a regional scope, whereas the temporal approach (annual surveys) is useful to determine the natural variability of these communities in order to discriminate between the observed changes due to natural factors and those related to the impact of disturbances related to global change (e.g. mass mortality outbreaks of invertebrates, invasive species, dramatic storms).

Contact address: `nteixido@icm.csic.es`