



HENNIG XL

ANNUAL MEETING OF THE WILLI HENNIG SOCIETY

CORNELL UNIVERSITY

ITHACA- USA

JULY 9-13, 2023

Cornell University is located on the traditional homelands of the Gayogóhó:nq' (the Cayuga Nation). The Gayogóhó:nq' are members of the Haudenosaunee Confederacy, an alliance of six sovereign Nations with a historic and contemporary presence on this land. The Confederacy precedes the establishment of Cornell University, New York state, and the United States of America. We acknowledge the painful history of Gayogóhó:nq' dispossession, and honor the ongoing connection of Gayogóhó:nq' people, past and present, to these lands and waters.

The XL Hennig Society Meeting is committed to creating a safe and harassment-free space of all the participants, regardless of race/ethnicity, gender identity and expression, sexual orientation, age, disability, and religion (or lack thereof). We are adopting a Presential and Virtual Code of Conduct based on the America Paleontological Society's Policy of Non-Discrimination and Code of Conduct (<https://www.paleosoc.org/non-discrimination-and-code-of-conduct>), which all participants are asked to agree to in order to register and participate.

ABSTRACTS

Abstracts are in alphabetical order by the surname of the first author.

ALMEIDA, EDUARDO A. B.¹, MARJORIE DA SILVA², ROGÉRIO B. LOPES^{1,2}, AND FERNANDO B. NOLL²

¹Dept. Biologia, Faculdade de Filosofia, Ciências e Letras, Universidade de São Paulo. Ribeirão Preto, Brazil; ²Dept. Zoologia e Botânica, Instituto de Biociências, Letras e Ciências Exatas (IBILCE), Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), São José do Rio Preto, Brazil. -Corresponding author: eduardoalmeida@usp.br

Evolution of neotropical swarm-founding wasps (Hymenoptera: Vespidae: Epiponini)

Neotropical swarm-founding wasps display extensive variation in several colony-level traits, making them an attractive model system for reconstructing the evolution of social phenotypes. Among these traits, caste dimorphism and variation in nest architecture are interesting examples of characteristics that can be investigated in a phylogenetic framework. Swarm-founding wasps are classified as Epiponini, a tribe of Politiinae (Vespidae) containing about 250 species classified within 19 genera. In this talk, we will present the phylogenetic results based on comprehensive taxon sampling and combining morphological, nesting, and molecular data. Molecular dating points to an early tribal diversification during the Eocene (ca. 55–38 million years ago), with the significant differentiation of current genera concentrated in the Oligocene/Miocene boundary. The resulting phylogenetic hypothesis impacts the interpretation of the evolution of certain behavioral traits, such as ovarian development, cast dimorphism, and aspects of the nest architecture.

ASHER, ROBERT

School of Biological Sciences, The University of Cambridge, Cambridge, UK. -Email: r.asher@zoo.cam.ac.uk

Fossilization and phylogenetic bias

I use the well-corroborated tree of living vertebrates to measure the phylogenetic value of data typically used in paleontology: bones and teeth. In particular, I ask if the loss of non-fossilizable data causes taxa to occur in misleadingly basal positions. Adding morphology to DNA data sets usually increases congruence of resulting topologies to the well-corroborated tree, but this varies among morphological data sets. Extant taxa with a high proportion of missing morphological characters can greatly reduce phylogenetic resolution when analyzed together with fossils. While loss of data can lead to misleadingly basal positions for fossils, there is no evidence that is more frequent than other kinds of phylogenetic error. Morphology comprises the evidence held in common by living taxa and most fossils, and phylogenetic analysis of fossils greatly benefits from inclusion of molecular and morphological data sampled for living taxa, whatever methods are used for phylogeny estimation.

BROWER, ANDREW V. Z.

USDA APHIS PPQ PEIP National Identification Services; AMNH Division of Invertebrates; Smithsonian NMNH Dept. of Entomology, USA. -Email: andrew.brower@usda.gov

30 years a cladist

Like the blind monks' descriptions of the elephant, there are undoubtedly as many versions of the history of cladistics as there are narrators. Some of these might be complementary; some might occupy "nonoverlapping magisteria"; some might be antithetical. My talk will touch on the theoretical issues that have interested me since I took up arms in 1993: ancestors, homology, branch length, accuracy, statistical inconsistency, the "assumptions" of cladistics, the nature of cladistic evidence. I can't promise to provide any new insights into this well-trodden ground, but I hope I can sketch a coherent picture of what cladistics means to me, how that epistemic framework is distinct and independent from evolutionary systematics, and why I think these perspectives are still worth arguing about.

CARPENTER, JAMES

American Museum of Natural History, 200 Central Park West, New York, NY 10024. USA. - Email: carpen@amnh.org

Cladistics at the American Museum of Natural History

The history of cladistics at the American Museum of Natural History is recounted.

ČERNÝ, DAVID

University of Chicago, Chicago, IL 60615, USA.- email: david.cerny1@gmail.com

Addressing heterotachy in morphological clock analyses using reversible-jump Markov chain Monte Carlo

In recent years, Bayesian tip-dating has seen increasingly widespread use in morphological phylogenetics owing to its ability to integrate character data and stratigraphic ages in a coherent statistical framework. An integral component of the tripartite model behind tip-dating is the (relaxed) morphological clock, which describes how the rate of character evolution varies across branches of the tree. Together with among-character rate heterogeneity models, relaxed clocks offer remarkable flexibility when it comes to modeling the evolution of discrete characters, but there remain scenarios which they cannot adequately capture. A basic example involves a situation in which branch A evolves faster than branch B for character x but slower than branch B for character y. This phenomenon, termed heterotachy and first described in the context of molecular evolution, has been suggested to be ubiquitous in discrete morphological data. Within the tip-dating framework, heterotachy can be addressed by partitioning characters based on similarities in their pattern of evolution, and assigning a separate clock model to each partition. However, since morphological datasets are minute in size compared to molecular alignments,

their ability to simultaneously estimate the huge number of branch rate parameters required by partitioned relaxed-clock models is questionable. Here, I extend an earlier method based on reversible-jump Markov chain Monte Carlo (rjMCMC) from branch lengths to branch rates in order to address this problem. Implemented in RevBayes v1.2, this approach allows each branch to sample only as many distinct rates as needed to explain the data. I apply the new model to a dataset of Mesozoic birds that was originally analyzed using a 6-partition unlinked clock model, and show that no branch favors more than four distinct rates. Using the R package NELSI, I also evaluate the ability of the model to recover the generating number of rates from simulated data. Together, these results suggest that sampling branch rates using rjMCMC makes it possible to accommodate heterotachy while avoiding the overparameterization inherent to fully partitioned analyses.

CZECH, LUCAS¹, JEFFREY SPENCE², YUNRU PENG¹, TATIANA BELLAGIO^{1,3}, XING WU¹, MEIXI LIN¹, AND MOISES EXPOSITO-ALONSO^{3,4}

¹Department of Plant Biology, Carnegie Institution for Science, Stanford, CA, USA;

²Department of Genetics, Stanford University, Stanford, CA, USA; ³Department of Biology, Stanford University, Stanford, CA, USA; ⁴Department of Global Ecology, Carnegie Institution for Science, Stanford, USA. -Corresponding author: lczech@carnegiescience.edu

Population genetic statistics for the next generation of pool sequencing

In recent decades, so-called Evolve-and-Resequence (EandR) experiments have become a popular approach to survey rapid evolution of populations over multiple generations. These experiments allow us to measure shifts in the allele frequencies of a population in response to new or shifting environmental conditions, such as a changing climate. Pool-sequencing of several individuals at once is a cost-effective and efficient tool to obtain reliable allele frequencies from a population of thousands to hundreds of thousands of individuals, and is often used in EandR experiments. However, specialized tools to efficiently analyze these data that take sampling noise stemming from the pool-sequencing approach into account were lacking. We developed two software tools to overcome statistical and bioinformatic challenges arising in this context. First, we present grenepipe, a workflow from raw sequencing data of individuals or pooled populations to genotypes (variant calling) and population allele frequencies. The pipeline automates trimming, mapping, variant calling, and quality control, with a selection of popular software tools in each of these steps, and produces variant calls and frequency tables. While generally applicable to individual sample data, it offers specialized steps for pool-sequencing. Our software downloads all dependencies and runs all steps automatically using a single command line call, and parallelizes processing for computer cluster environments, allowing large datasets to be analyzed efficiently. Second, to enable inferences of evolutionary signatures from frequency data, we created grenedalf, a C++ command line tool to compute population genetic statistics. It computes unbiased statistics of F_{st} , Π , Tajima's D with pool-sequencing data, far outperforming alternative tools, and fixing long-standing issues in the existing tools. Further it offers novel data exploration tools such as windowed allele frequency spectrum visualizations based on the allele frequencies, and built-in data filters and manipulations. Apart from its speed, it offers many convenience options, such as reading all standard file formats. These tools are designed for scalability and ease-of-use with contemporary file formats, which we showcase

using the GrENE-net.org project, a large-scale Evolve-and-Resequencing experiment with *Arabidopsis thaliana* from across the world."

CZECH, LUCAS¹, AND ALEXANDROS STAMATAKIS^{2,3}

¹Department of Plant Biology, Carnegie Institution for Science, Stanford, CA, United States;

²Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies; Heidelberg, Germany; ³Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany. -Corresponding author: lczech@carnegiescience.edu

Leveraging phylogenetic placement to understand the environmental drivers of microbial community composition

In recent years, phylogenetic placement has emerged as a powerful tool for investigating the evolutionary relationships and dynamics of microbial communities. By assigning (microbial) sequences to their most likely positions within a phylogenetic reference tree, phylogenetic placement enables us to gain valuable insights, for instance, into the origins, transmission, and evolution of infectious agents. This talk explores several methods for the analysis and visualization of phylogenetic placements, specifically focusing on microbial and meta-genomic samples, and highlighting potential applications to infectious diseases research. In particular, this talk will go beyond employing phylogenetic placement for traditional sequence classification. Various placement-based methodologies have been developed to infer the influential factors driving the composition of microbial communities within a phylogenetic framework, such as Edge PCA, Edge Correlation, and Placement-Factorization. These methods enable us to identify key branches in the tree, thus shedding light on the evolutionary dynamics and traits associated with a set of microbial samples. This approach allows us to unravel the interplay between the microbial abundances, host factors, and environmental variables, and can potentially facilitate an understanding of microbial communities in their evolutionary context. This talk aims to provide an overview of these advancements in phylogenetic placement methods and their applications.

DIKOW, TORSTEN

Smithsonian National Museum of Natural History, Washington, DC., USA. -Email:

DikowT@si.edu

New discoveries and phylogenetic placement of assassin flies (Insecta: Diptera: Asilidae) in Cretaceous and Paleogene ambers

The extinct assassin flies (Asilidae) preserved in Cretaceous and Paleogene ambers are reviewed and phylogenetically placed. The fauna in Cretaceous ambers from Myanmar and the USA (Burmese and Raritan) have recently been reviewed recognizing one species from each amber deposit based on three and one specimen, respectively. An additional female specimen preserved in Burmese Amber is described representing a new species, which does not belong in the recently established genus †Burmepogon. Paleogene: Eocene amber from northern Europe (Baltic, Bitterfeld, and Rovno) is well-known for preserving insects. Three Asilidae species have

been recorded from Baltic and Bitterfeld ambers to date. With the availability of 28 additional pieces of Baltic and Rovno amber with assassin-fly inclusions, the species diversity can be increased to six, representing two of the previously described species. Several inclusions are not well-preserved to allow formal description and placement. It is concluded that the Baltic, Bitterfeld, and Rovno deposits preserve the same insect fauna and that the most abundant Asilidae in Eocene amber, *Protoloewinella keilbachi* (Laphriinae, 16 specimens), is found in all three deposits. Five genera and four species are hypothesized to represent new, undescribed taxa. The phylogenetic placement is based on an expanded morphological matrix based on Dikow (2009, 220 characters) representing all subfamily taxa, additional extant species, and nine amber fossils. The Cretaceous amber species could not be coded for 32–78% and Paleogene amber species for 28–48% of characters. The backbone relationships within Asilidae remain largely unresolved but several subfamily taxa are recovered as monophyletic. Several Eocene amber fossil taxa enhance the relationships within these subfamily taxa and are grouped deeply within radiations. The unique character combinations found in some extinct taxa make phylogenetic placement a challenge.

DORNBURG, ALEX¹, HAYLEY B. HASSLER², AND JEFFREY P. TOWNSEND³

¹Department of Bioinformatics and Genomics, University of North Carolina, Charlotte, North Carolina 28223, USA; ²Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, GA 30332, USA; ³Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut 06510, USA. -Corresponding author: adornbur@uncc.edu

An evolving future for pandemic preparedness: evolutionary medicine and phylogenetic comparative methods inform the durability of SARS-CoV-2 immunity

Over the next century, human population growth and consequent interactions with wildlife are expected to escalate the frequency of viral zoonotic transmission. Consequently, proactive measures preparing for the emergence of novel viral pathogens have become prudent. Effective combat against a novel virus is hindered by a lack of knowledge regarding the duration of protective immunity—information that is essential to modeling and dictates both short- and long-term policy responses. Estimation of the durability of immunity would classically be presumed to require multi-year, long-term studies. However, the COVID-19 pandemic highlighted the pyrrhic consequences of waiting out the extensive timeline required for traditional data collection to obtain vital policy-relevant information. This temporal mismatch underscores the value of rapid approaches that would bridge this knowledge gap in pandemic preparedness by estimating likely times to reinfection and the consequent durability of immunity. Here we demonstrate the robust viability of leveraging evolutionary approaches with long-term infection and antibody data in response to current human-infecting or animal-infecting viruses. Using this approach we estimate infection probabilities in relation to antibody declines and translate these probabilities into inferred times to reinfection and durabilities of immunity against SARS-CoV-2. We further demonstrate that this approach can be extended to estimate distribution of likely times to breakthrough infection following vaccination against SARS-CoV-2 as well as the probability of breakthrough infection under alternate mRNA booster vaccination schedules for immuno-typical individuals as well as those undergoing various cancer treatments.

These results demonstrate the utility of comparative phylogenetic methods in providing a quantitative basis for otherwise unknown parameters that are fundamental to personal, clinical, and public health policy decision-making.

EZCURRA, MARTÍN D.

Sección Paleontología de Vertebrados, CONICET–Museo Argentino de Ciencias Naturales (MACN). Ángel Gallardo 470, C1405DJR Ciudad Autónoma de Buenos Aires, Argentina. - Email: martindezcurra@yahoo.com.ar

Exploring different weightings against homoplasy in genealogies of paleontological phylogenetic matrices

Cladistics is the most widespread methodology to reconstruct the phylogenetic relationships among species in paleontology and character weighting has been one of its most controversial issues during decades. Although simulations have shown that implied weights (a way to weight characters against homoplasy) outperform equal weights, weighting against homoplasy currently lack an extended usage in paleontology. Iterative modifications of several phylogenetic matrices in the last decades resulted in genealogies of datasets that allow evaluating the effect of different character weightings against homoplasy directly in empirical data. Each generation was compared against the most recent generation in each genealogy because it is here assumed that the latter has the most comprehensive (taxon and character sampling) and complete (lower amount of missing data for taxa present in previous generations) matrix and, thus, it is the most reliable version of the matrix at that moment. The analyses were conducted in five different direct genealogical lines composed of eight to thirteen genealogies. Each generation was analyzed under equal weightings (EW) and implied weightings (IW) and extended implied weightings (EIW) with a range of concavity constant values (k) between 3–30. Pairwise comparisons between trees were conducted using Robinson–Foulds distances normalized by total number of groups, distortion coefficients, subtree pruning and regrafting movements, and proportion of different groups. The results were consistent in all genealogies, in which IW and EIW outperformed EW in the vast majority of cases for all comparative measures (with the exception of very low k values; e.g., $k = 3–5$). The results under IW and EIW showed very minor differences between them. A Principal Components Analysis based on the comparative measures was used to calculate Euclidean distances between each character treatment (k values and EW) and the latest generation. The different genealogies showed different ranges of k values that retrieved the most similar results to the latest genealogy. There was a significant positive linear correlation between the optimal k values and the number of terminals of the latest generations. This result is promising to try to establish a generalization between matrix size and an optimal range of k values.

GARZÓN-ORDUÑA, IVONNE J.¹, AMBROSIO TORRES G.², AND ANDREW V.Z. BROWER³

¹Instituto de Biología, Universidad Nacional Autónoma de México; ²Center for Integrative Biodiversity Discovery (Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science), Berlin, 10115 Germany. ³USDA APHIS PPQ PEIP National

Identification Services; AMNH Division of Invertebrates; Smithsonian NMNH Dept. of Entomology, USA. -Corresponding author: ivonne.garzon@ib.unam.mx

Patterns of missing data and completeness in UCE datasets

We explore the relationship between “Completeness” and missing data (MD) (numbers of nucleotides missing per taxon, per loci, or overall) in UCE datasets (matrices of Ultraconserved Elements), and based on this assessment determined whether or not “Completeness” by itself should continue to be used as a surrogate for the amount of MD. Authors that use UCE, built their datasets using Phyluce, a software package that includes tools that assist the analytical steps involved in the transformation of the data from raw, unassembled sequences to the alignment stage. Phyluce counts the number of taxa for which data was successfully generated for each UCE locus, identifies these loci according to the user’s preferred threshold of completeness, and generates a matrix. For example, a matrix generated under a “100% completeness” preference, would include only those loci for which all the sampled taxa have data. Missing data per se have rarely been addressed by users of UCE data. Instead and most commonly, authors refer to the completeness percentage used as an indirect approximation, or refer to other statistics of their datasets, such as number of loci, number of nucleotides, and number of phylogenetically informative characters. Moreover, there appears to be no standard or preferred completeness value: some authors have chosen a 90% completeness matrix, others, chose matrices with 75%, 70%, 50% and even 25% completeness. We used multiple UCE datasets to examine the patterns of MD and its relationship to Completeness. We found that MD does not linearly track the completeness measure as one would expect. In other words, a matrix built under 90% Completeness misses more than 10% of its data and in multiple dimensions (per taxa and per loci). Although higher completeness percentages do accomplish the generation of matrices whose terminals have less amount of missing data on average (and as a consequence, matrices with less missing data overall), it does so unsatisfactorily and inaccurately. Overall, a high completeness value does not prevent the inclusion of loci with many missing entries.

GOLOBOFF, PABLO¹ AND JAN DE LAET²

¹Unidad Ejecutora Lillo, UEL CONICET – Fundación Miguel Lillo, Tucumán Argentina;

²Göteborgs Botaniska Trädgård, Göteborg, Sweden. -Corresponding author: pablogolo@yahoo.com.ar

Farewell to the need for character independence: phylogenetic methods to incorporate different types of dependence between characters

This paper discusses methods to take into account interactions between characters, in the context of parsimony analysis. These interactions can be in the form of some characters becoming inapplicable given certain states of (an)other, primary character(s); in the form of only certain states being allowed in some character(s) when (an)other character(s) have (has) a given state or set of states; or in the form of transformation costs in some character being higher or lower when (an)other character(s) have (has) given states or transformations between states. Character-state reconstructions and evaluation of trees under the assumption of independence may easily lead to

ancestral assignments that violate elementary rules of biomechanics, well-established theories relating form and function, or ideas about character co-variation. An obvious example is reconstructing an ancestral bird as wingless and flying at the same time; another is reconstructing a protein-coding gene as having a stop codon in some ancestors. If the characters are optimized independently, such chimeric ancestral reconstructions can occur even when no terminal displays the impossible combination of states. A set of conventions (implemented via new TNT commands and options) allows defining complex rules of interaction. By recoding groups of characters with proper step–matrix costs (and excluding impossible combinations from the set of permissible states), it is possible to find the ancestral reconstructions that maximize homology (and thus the degree to which similarities can be explained by common ancestry), within the constraints imposed by the rules specified by the user. We expect that considerations of biomechanics, functional morphology, and natural history, will be a source of many theories on possible character dependences, and that the present implementation will encourage users to take the possibility of character dependences into account in their phylogenetic analyses

GRAMS, MARKUS AND STEFAN RICHTER

University of Rostock, Mecklenburg-Vorpommern, Germany. -Corresponding author:
markus.mendes-grams@uni-rostock.de

Malacostracan phylogeny and the importance of morphological character conceptualization

Morphological matrices are still a valuable and necessary tool for cladistic analyses. They not only allow to include taxa still eluding genetic examination (e.g. fossils), but also hold explanatory power through offering evolutionary narratives by outlining phenotypic transformations. Although often seen only as numerical summaries of observations for the purpose of cladistic analyses, morphological matrices also hold value as collections of ideas, concepts and the current state of knowledge. An important and actually most central step is the intellectual process of character conceptualization, which a priori requires a deep understanding of the to-be characterized features (morphemes) and their properties throughout the taxon sampling. While this includes such obvious steps as establishing homology hypotheses of morphemes as characters and deciding which differences are conceptualized as (how many) states, it also involves the identification of ontological character dependencies and decisions on character weighting and the polarization between states. To explain important nuances in character conceptualization, malacostracan morphology is used as a practical example. Malacostracan phylogeny has remained debated in many aspects. Morphological and molecular approaches have proposed as also challenged various monophyletic groups, with only few being currently accepted with little to no doubt. Be it the validity of the Eucarida (Decapoda + Euphausiacea), the positions of Amphionidacea and Thermosbaenacea, the affiliation of Bathynellacea to Anaspidacea (as Syncarida) or the monophyly of Mysidacea or their belonging to Peracarida (not to mention more specific relationships within that taxon) – many phylogenetic questions could not be answered satisfactorily up to today. A collection and careful revision of established character concepts, complemented by new ones, resulted in 201 morphological characters for 34 terminal taxa, representing all major lineages of the Malacostraca, including many of the more puzzling taxa such as Amphionidacea, Procarididae, Bathynellacea,

Stygiomysida and Thermosbaenacea. For the cladistic analysis we tested different ways of conceptualization for critical character complexes, including different methodological approaches, such as implied character weighting and the application of step matrices. Based on the phylogenetic results problematic ways of character conceptualization and other analytical issues are identified and discussed.

HARTOP, EMILY

Center for Integrative Biodiversity Discovery, Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Invalidenstr. 43, 10115 Berlin, Germany. -Email: emily.a.hartop@gmail.com

The how and why of tackling a genus of flies that has 100,000 undescribed species

The taxonomic problems faced in hyperdiverse groups are not merely larger versions of the problems encountered in other taxa. Rather, the scale and complexity of these groups, combined with our ignorance of them, necessitates new and innovative solutions. How can we identify, organise, and classify groups with thousands of unknown species, countless numbers of individuals, challenging morphology coupled with very small size, scant life history data, troublesome trees, and species curves that refuse to plateau no matter how much sampling you do? This talk will focus on the arsenal of approaches that have been applied to mega-genus *Megaselia* in recent years, summarising an (ever evolving) approach to a dark taxon. Numerous efforts are contributing data to the comprehensive understand of this genus, ranging from systematic and integrative approaches to species discovery, delimitation and description, spatial and temporal analyses, ecological modelling, experimental taxon sampling, microbiome analysis, morphometrics, convolutional neural networks and robotics, and advanced morphological scans. All of these pieces contribute to a better understanding of one of the largest genera of animals. What is next and how do we bring all of this information together?

HÖPEL, C.G., SCHWENTNER, M., AND RICHTER, S.

Allgemeine und Spezielle Zoologie, Universität Rostock, Germany. -Corresponding author: christoph.hoepel@uni-rostock.de

*Speciation in correlation with cave immigration in Tasmanian mountain shrimps
(Anaspides, Malacostraca, Crustacea)*

The processes and underlying mechanisms that result in the colonization of extreme habitats such as caves and groundwater, are among the most fascinating topics in evolutionary biology. Special adaptations are necessary for surviving in caves and other subterranean habitats. Permanent darkness and low food availability present major challenges for new colonizers. Studying cave biodiversity allows answering questions about the interaction of adaptation, phylogeography and evolution. The Tasmanian Mountain Shrimps (*Anaspides* spp., Anaspidae, Malacostraca) offer the unique possibility to study colonization and adaptation to a cave or subterranean habitat in closely related but separate lineages. Our phylogenetic analyses show that within the genus *Anaspides*, colonization occurred many times independently and at

different time levels. We also studied the genetic differentiation within multiple instances of cave immigration and their influence on the speciation. Genetic differentiation was studied using Sanger Sequencing and ddRAD. One of our main focus areas is the Mt Field National Park, where two closely related species with several lineages can be found. While tarns and pools on the mountain plateau are inhabited by *A. richardsoni*, the mountain range surrounding caves are inhabited by its sister species, the obligate cave dwelling *A. eberhardi*. Our analyses indicate multiple instances of immigration into the subterranean habitat of the ancestors of what is today *A. eberhardi*, which appear to have occurred during different time periods of the Pleistocene. Interestingly the inflow streams to these caves are inhabited only by distant related species. We found evidence of interbreeding among populations of *A. eberhardi* and the absence of interspecific interbreeding between *A. eberhardi* and *A. richardsoni*, which indicates the presence of two well defined species according the Hennigian species concept despite the independent and potential polyphyletic origin of *A. eberhardi*.

HOYAL-CUTHILL, JENNIFER¹ AND NICHOLAS GUTTENBERG²

¹School of Life Sciences, University of Essex; ²Cross Labs, Cross Compass Ltd., Kyoto, Japan. - Corresponding author: j.hoyal-cuthill@essex.ac.uk

New machine learning methods for inferring phylogenies from embeddings of phenotypic photographs

Morphological character analysis by a human systematist has previously been an essential first step for most morphological phylogenetics. However, this subjective process is often undocumented and has been previously shown to be non-replicable in tested cases. Consequently, it could be argued that morphological systematics has hitherto rested on a fundamentally unscientific footing. New machine learning (ML) methods enable automated and replicable reconstruction of morphological phylogenies from phenotypic photographs, without the need for human morphological character analysis. Furthermore, machine learning methods are capable of simultaneously analysing datasets covering thousands of biological specimens, or more. However, new applications of machine learning for phenomic phylogenetic reconstruction raise many methodological options to test, such as the choice and performance of different ML methods, network architectures, parameters and data sampling protocols. One machine learning method which has shown promise for phenomic phylogenetics is supervised triplet embedding. Triplet embedding methods train a convolutional neural network, which analyses features of input images, to “embed” or locate each image in a multidimensional space in which proximity represents image similarity, for example by Euclidean distance. This results in an embedding in which the distance between images is a Euclidean, metric representation of their phenotypic similarity, as learnt directly from the input images. An output is a distance matrix of image distances which can be used as the input for phylogenetic analysis, for example, most simply, by neighbour joining of taxa based on the average distance between their representative images, though a range of other phylogenetic methods can be potentially applied to the image embedding locations. Our applications of triplet embeddings provide case studies from whole-specimen photographs of butterfly museum specimens. First, dense taxonomic sampling that provides links between included taxa may improve the recovery of phylogenetic signal by machine learning (measured in comparison with phylogenetic signals from analytically independent genetic data

for overlapping taxa). Second, ML training time may affect the strength and depth of phylogenetic signal recovered. Overall, ML triplet embedding can be used to reconstruct inter-taxon phenotypic distances, automatically from phenotypic images, that show significant correlations to genetic distances, offering considerable potential for the field of systematics.

JANIES, DANIEL

University of North Carolina at Charlotte, USA. -Email: djanies@uncc.edu

Phylogenetics as applied to infectious diseases

In this talk I shall introduce the symposium "Phylogenetics as applied to infectious diseases". I shall also discuss the development of the Center for Computational Intelligence to Predict Health and Environmental Risks (CIPHER) at the University of North Carolina at Charlotte. I shall illustrate how we use phylogenetics to address folklore in virology.

LAURIN, MICHEL¹ AND GILLES DIDIER²

¹ CNRS/MNHN/UPMC, Sorbonne Université, Muséum National d'Histoire Naturelle, CR2P ("Centre de Recherches sur la Paléobiodiversité et les Paléoenvironnements" UMR 7207), Paris, France; ² Univ Montpellier, CNRS, IMAG, Montpellier, France. -Corresponding author: michel.laurin@mnhn.fr

How did Ophiacodontidae, Edaphosauridae and Sphenacodontidae become extinct? New insights provided by a FBD model with variable rates

Synapsid evolution is marked by two successive evolutionary radiations, in the late Carboniferous and in the Middle Permian, respectively. The first of these gave rise to taxa such as Ophiacodontidae, Edaphosauridae and Sphenacodontidae, which were the dominant top predators and some of the most common herbivores in the Cisuralian. The extinction of these taxa remains enigmatic; some authors have postulated that a major biological crisis caused their extinction, but our recent study, published in 2021, finds a gradual decline that is not typical of mass extinction events. To better understand the demise of these emblematic taxa, we study an updated dataset (with age assignments reflecting recent literature) using a piecewise-constant implementation of the FBD model that allows the three rates (cladogenesis, extinction and fossilization) to vary between pre-defined time slices (users may specify constraints to force some of the rates to remain the same in some of the slices). This implementation also allows to model massive extinction events by sampling the lineages extant at the end of a slice. Rate estimation and model choice are conducted through a MCMC framework.

LEE, LESHON¹, BRIAN BROWN², RUDOLF MEIER³, ANG YUCHEN⁴, AND EMILY HARTOP³

¹National University of Singapore, Department of Biological Sciences, 16 Science Drive 4, Singapore 117558; ² Natural History Museum of Los Angeles County, 900 Exposition Blvd, Los Angeles, California 90007, United States; ³ Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Center for Integrative Biodiversity Discovery,

Invalidenstraße 43, 10115 Berlin, Germany; ⁴ National University of Singapore, Lee Kong Chian Natural History Museum, 2 Conservatory Dr, Singapore 117377. -Corresponding author: leshonlee@u.nus.edu

Experimental Taxonomy for Dark Taxa: Addressing taxonomic impediments with “divide and conquer” and prudent disregard

Most of the world’s animal biodiversity remains undescribed and is concentrated in “dark taxa” characterized by few described species in an ocean of undescribed diversity. This interferes with biodiversity discovery, inventories, regular monitoring, and filing new information under scientific names in the biological literature. Fortunately, recent advances in the sorting of specimens to putative species with molecular markers have greatly sped up the discovery process. However, two challenges remain. The first is efficient methods for integrative species description, which can address the “dark taxon impediment”. The second is distinguishing between described and new species to address the “superficial description impediment”. Here we address both challenges by first testing whether “divide and conquer” could be applied to dark taxa. We start by using integrative taxonomy to delimit species of Phoridae based on a minimalist sample; i.e., specimens from only one two-week Malaise trap sample. We find a surprisingly high number of 28 phorid species. We then test whether the species remain stable with increased spatio-temporal sampling. We demonstrate that 27 of the 28 species remain distinct even after examining more samples from the same trap, sampling many additional localities in Singapore and testing the species with material from Indonesia. Only one species, *Megaselia hwangi*, failed the test because it was initially split into three species based on the two week sample. Overall, our findings thus suggest that dark taxa can be tackled by applying divide-and-conquer to small, local samples. Next we address the “superficial description impediment” by surveying the existing taxonomic literature for Indo-malayan phorids. Most names have only been used once and we argue that not all names should be resolved before describing the diversity. We thus go ahead and describe 69 new species collected by a single Malaise trap placed in secondary forest on Sentosa Island. We suggest how an efficient workflow illustrating how dark taxa could be described.

LLOYD, GRAEME¹ AND JENNIFER HOYAL CUTHILL²

¹ School of Earth and Environment, University of Leeds, Leeds, UK; ² School of Life Sciences, University of Essex, UK. -Corresponding author: graemetlloyd@gmail.com

Limits and proofs under parsimony across the full empirical and theoretical range of discrete cladistic characters

Quantifying limits (m = minimum cost, g = maximum cost, g_{max} = maximum possible cost) under parsimony are critical for measuring homoplasy. However, to date these limits have only really been considered for a limited range of character types meaning they are not all underpinned by mathematical proofs. After examining 4,467 primarily morphological cladistic data sets we define twelve distinct discrete character types that span the full range of both empirical use and theoretical possibility. These can be further aggregated into a single general case (a cost matrix) and one special case (Dollo). We further collect seven rules that define a mathematically valid

cost matrix and review existing limits and proofs. We show that m is generally solvable by taking the length of the minimum spanning tree of the state graph implied by a cost matrix. We show that an existing proof that the star tree is the longest possible tree solves g in general (i.e., for any cost matrix). We then show that generally g can be maximised for the unique state frequency across t tips where any ancestral state of the star tree is equally parsimonious. Finally, we show that Dollo is a special case requiring unique solutions to define the three limits. We conclude by considering avenues for future work that are key to converting these theoretical limits into practical ones.

MACHADO, DENIS JACOB

University of North Carolina at Charlotte, College of Computing and Informatics, Department of Bioinformatics and Genomics; University of North Carolina at Charlotte, CIPHER center, USA.
-Email: dmachado@charlotte.edu

Combining phylogenetic and artificial intelligence to unravel viral recombination

Recombination is key to unlocking the intricate evolution of viruses, fueling their genetic diversity, shaping their virulence, and enabling their cunning escape from immune responses. Astonishingly, more than 35 cutting-edge computational algorithms have emerged, employing advanced phylogenetic methods, site patterns analysis, pairwise sequence comparisons, and population genetics approaches to detect recombination events in viral sequences. However, the demand for substantial computational resources or limitations in handling vast genomic databases have hindered their progress. Moreover, these methodologies have often overlooked the potential impact of evolutionary convergence. This talk delves into the current state-of-the-art approaches for estimating recombination in viruses, shedding light on their triumphs and pitfalls. It also introduces an innovative methodology that melds the powers of phylogenetics and artificial intelligence to unravel viral recombination. This groundbreaking technique distinguishes evolutionary convergence from genetic recombination and shows exceptional promise in revolutionizing our understanding of viral dynamics.

MORRONE, JUAN J.

Museo de Zoología ‘Alfonso L. Herrera’, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México (UNAM), 04510 Mexico City, Mexico. - Email: juanmorrone2001@yahoo.com.mx

History of phylogenetic/cladistic biogeography

Willi Hennig (1960) and Lars Brundin (1966) applied an explicit phylogenetic approach to study the geographical distribution of life from a historical viewpoint. This approach, known as phylogenetic biogeography, uses phylogenetic trees to infer dispersal, applying the progression and deviation rules, and allows to infer vicariance patterns. In the 1970s and 1980s, Gareth Nelson, Don E. Rosen and Norman Platnick associated Léon Croizat’s panbiogeography with Hennig’s phylogenetic systematics, creating cladistic biogeography, that is a strictly vicariance approach. I have concluded previously that both phylogenetic and cladistic biogeography have a

place in contemporary biogeography, the former for analysing the biogeographic history of particular taxa (taxon biogeography), and the latter when searching for general, vicariance patterns for different plant and animal taxa inhabiting the same areas (biota biogeography). I also noted that the use of “phylogenetic biogeography” for some methods of cladistic biogeography is incorrect and should be avoided. I would like to propose that cladistic biogeography is a valid approach for biogeography, whereas phylogenetic biogeography should have a special place within a more integrative systematics, that goes beyond narrow conceptions of the discipline, within a truly Hennigian perspective.

NIXON, KEVIN C.

L.H. Bailey Hortorium, Plant Biology Section, SIPS, Cornell University, Ithaca, NY, USA. -
Email: kcn2@cornell.edu

The evolution of phylogenetic methodology in relation to modern concepts of homology.

Beginning in the mid 19th Century, a strong interest in determining phylogenetic relationships among both living and fossil organisms began, particularly following the groundbreaking works of Darwin, and followed by numerous workers who floundered without clear and repeatable methodologies. These efforts accelerated with an increased understanding of genetics in the early 20th Century, but still was held back by the lack of a clear methodology to reconstruct phylogenetic trees. This changed dramatically with the introduction of basic modern phylogenetic concepts by Hennig, but methods were still idiosyncratic and vague in application. The landscape was further complicated by the competition between purely phenetic approaches and truly Hennigian approaches until the latter eventually prevailed, particularly with the promotion of these methods by several workers at the American Museum of Natural History. However, it was not until Farris introduced tree-building methods and clarified many of the important phylogenetic concepts in papers from the late 1960s through the 1980s, accompanied by many others, that a robust and repeatable way of reconstructing evolutionary history emerged, based largely on the concept of parsimony, as a way to resolve conflict in presumed homology. Farris’s tree building algorithms, which allowed evaluation of massive numbers of possible trees, were modified and later improved by others such as Goloboff, and together these became the basis for almost all modern phylogenetic programs, whether utilizing parsimony, likelihood or Bayesian criteria. Viewing these scholarly events and transitions in the context of homology, and how homology is related to each of the major tree construction methods, can provide further insight into the evolution of the field of phylogenetics.

NOLL, FERNANDO B.¹, EDUARDO A. B. ALMEIDA², LAURA CHAVARRIA-PIZZARRO³, AND MARJORIE DA SILVA¹

¹Departamento de Ciências Biológicas, IBILCE-UNESP, Rua Cristovão Colombo 2265, CEP 15055-240 São Jose do Rio Preto, Brazil; ²Departamento de Biologia, FFCLRP-USP, Avenida Bandeirantes 3900, CEP 14040-901 Ribeirão Preto, Brazil; ³Instituto Tecnológico de Costa Rica, Escuela de Biología-Centro de Investigación en Biotecnología, Cartago, Costa Rica. -

Corresponding author: fernando.noll@unesp.br

Untangling concepts for social wasps: how phylogenies help us understand the evolution of sociality in Neotropical swarm founding wasps

Neotropical swarm founding wasps (Epiponini) are polygynic, with multiple queens alternating over the colony cycle. There are several potential queens in the early stages of this cycle, but as it progresses, the number of queens is reduced. Because most individuals remain reproductively totipotent, there is great potential for conflicts over reproduction. Workers could have an advantage in controlling queen production because they are much more numerous than queens. Nevertheless, the queen selection process is little known for these wasps. We aimed to study the behavior of queens and workers during queen selection in multiple species of Epiponini, integrate information from previous behavioral studies, and perform a comparative analysis to interpret changes evolutionarily. We conducted observations on nine species belonging to five genera: *Brachygastra*, *Chartergellus*, *Metapolybia*, *Polybia* and *Protopolybia*. Females were individually marked to make direct and video observations. Queen production was artificially induced. A total of 28 behaviours related to queen selection were identified. The most aggressive interactions between castes, such as bite and dart, were lost in the major lineages of Epiponini. Bending display I is an ancient behaviour used as the main dominance display. Behaviours exhibited by workers to test queen status arose in the common ancestor of the Epiponini and are not shared by other polistine wasps. Consequently, the act of workers testing queen status probably was present in the Epiponini ancestor. Ritualized test display and dominance behaviors are used in Epiponini as honest signals of the queen's reproductive potential instead of aggressive behaviors. Caste flexibility had already been suggested as the ground plan for Epiponini and is herein discussed as decisive for colony survival of swarm wasps, because it allows colonies to respond efficiently to different situations that may eventually arise.

OCAÑA, KARY, LUCAS SILVA, MICAELLA COELHO, RAFAEL TERRA, AND CARLA OSTHOFF
National Laboratory of Scientific Computing (LNCC) of the Brazilian Ministry of Science, Technology and Innovation (MCTI). -Corresponding author: karyann@lncc.br

Phylogenetic and evolutionary analysis of dengue virus serotypes in Brazil: User-steering in large-scale phylogenomic workflows

Scientific workflows that construct phylogenetic trees have been proposed to manage the complex interplay of multiple bioinformatics applications. The exploratory nature of phylogenetic analyses is considered a computing-intensive process and demands monitoring high-performance computing (HPC) executions. This talk describes the development of a workflow implementation named PhyloHPC which allows for the remote execution of pipelines for phylogenetic analyses while benefiting from the parallel processing power of HPC environments. PhyloHPC is coupled to the Parsl Python library which contributes to the management of tasks executions in a HPC environment. Computational speedup and total execution time of the workflow show significant performance improvements and demonstrate that the Parsl model definition brings advantages of large-scale parallelism and the provenance support allows for runtime monitoring. We also explore the performance and scalability of RAXML in HPC environments. Phylogenetic tools included in PhyloHPC cover a wide yet selected range of programs and packages such as PHYLIP, PhyML, Weighbor, Tree-Puzzle,

RAxML, Garli, IQ-TREE, and MrBayes. RAxML is a popular maximum likelihood software for processing large amounts of data in HPC environments. RAxML implements several phylogenetic likelihood function kernel variants (SSE3, AVX, AVX2) and offers coarse-grain/fine-grain parallelism via Hybrid and MPI/PThread versions. The circulation of new dengue virus lineages was described in Brazil and the Americas, but the precise classification of genotypes and lineages (clades) has been controversial. Here we perform phylogenetic and nucleotide-distance analyses of high-quality complete DENV1-4 genomes. DENV1 genotype V is representative of Latino America, and isolates in Brazil were organized into three distinct lineages, named clades I, II, and III. Sequence alignment was performed using MAFFT and phylogenetic trees were estimated using RAxML under the GTR nucleotide substitution model, which was inferred as the best-fit model by ModelFinder. The robustness of the tree topology was determined using 1,000 bootstrap replicates and visualized using FigTree v1.4.4. Due to the connectivity between the potential circulation of arboviruses, this investigation will be of value to scientists and public health officials dealing with the control and prevention of infections by this arboviral pathogen.

PITTMAN, MICHAEL

School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China. -Email: mpittman@cuhk.edu.hk

The origin of flight in theropod dinosaurs: new insights from methods using palaeontological & morphological datasets

Recent advances in our understanding of the origins of flight in theropods have greatly benefited from ongoing progress in reconstructing the phylogeny of feathered dinosaurs and early birds, as well as a range of methods using palaeontological and morphological datasets. Ancestral state reconstruction analysis of proxies for powered flight potential (e.g. specific lift and wing loading) have confirmed powered flight capabilities among early birds. These analyses have also revealed the presence of powered flight in select close-related theropods (e.g. dromaeosaurid *Microraptor*) that make a compelling case for multiple originations of flight among dinosaurs. Did heterochrony play a role in flight origins? Using ancestral landmark-based shape ontogenies of theropod arms/wings that account for changes in developmental timing, this can now be tested. New approaches to analysing characters with partial dependences resulting from functional or biomechanical constraints are promising. Could they help to resolve long-standing phylogenetic uncertainties? Could they encourage development of flight-related evo-devo hypotheses? This talk concludes with priority areas for future research that aim to address key remaining questions in the study of theropod flight origins.

POL, DIEGO¹, ALAN TURNER², AND ERIC WILBERG²

¹CONICET, Museo Paleontológico Egidio Feruglio, Trelew, Chubut, Argentina; ²State University of New York NY Stony Brook, Stony Brook, NY, USA. -Corresponding author: dpol@mef.org.ar

Taxon instability in the phylogeny of Crocodyliformes: evaluating causes and impact on low support

Crocodyliformes is one of the two major clades of Archosaur reptiles that survive until today. Although their extant diversity is small, in the Mesozoic the group experienced a remarkable radiation into terrestrial, freshwater, and marine environments. Here, we explore a new global dataset of Crocodyliformes and address the identification of unstable taxa and clades, highlighting those that decrease consensus resolution and support measures in parsimony analysis. The taxa identified are evaluated to determine the underlying causes of their instability, such as missing data (lack of information) or character conflict. Abundant missing entries are common in paleontological datasets and their presence is the most common cause of instability for the taxa or clades that decrease support values and consensus resolution. However, two clades, Sebecidae and Thalattosuchia, are identified as clustering large amount of homoplasy that underlies their alternative positions in trees analyzed during the nodal support evaluation.

RICHTER, STEFAN

Department of Zoology, University of Rostock, Germany. -Email: stefan.richter@uni-rostock.de

Phylogenetic systematics vs. cladistics - a German perspective

Depending on the authority, cladistics is considered either the "contemporary version of Hennigian phylogenetic systematics" or "a form of non-phylogenetic systematics" or "defined by its application (Wagner parsimony)." The general principles of (traditional) Hennigian phylogenetic systematics are outlined, especially the role of parsimony. The major difference between the more traditional Hennigian phylogenetic systematics that prevailed in Germany in the last decades of the 20th century and (modern-day) cladistics was the result of the increasing use of numerical techniques for the phylogeny reconstruction in contemporary cladistics. Some important events of this "clash of civilizations" in the last century are described and conceptual differences are discussed.

SILVIRIA, JACQUELINE S.¹, CHRISTINA I. BARRÓN-ORTIZ², AND GREGORY P. WILSON MANTILLA³

¹ Department of Earth and Space Science, University of Washington, Seattle, Washington, USA; Burke Museum of Natural History and Culture, University of Washington, Seattle, Washington, USA; ² Quaternary Paleontology Program, Royal Alberta Museum, Edmonton, Alberta, Canada

³ Department of Earth and Planetary Science, University of Washington, Seattle, Washington, USA; Burke Museum of Natural History and Culture, University of Washington, Seattle, Washington, USA; Department of Biology, University of Washington, Seattle, Washington, USA. -Corresponding author: jsilvi@uw.edu

Specimen-level cladistic analyses of geometric morphometric configurations from mammalian postcanine dentition: Getting the state of the art into shape

Specimen-level studies of mammalian postcanine dentition commonly employ non-phylogenetic analyses of geometric morphometric (GM) occlusal configurations, such as principal component analysis (PCA). Rarely are dental GM data analyzed in a phylogenetic context. Here we present results from experimental maximum parsimony cladistic analyses in TNT 1.6, using two datasets of 2D mammalian occlusal configurations: 186 specimens (113 M2, 73 m2) from 36 species of latest Cretaceous-earliest Paleogene North American eutherians and metatherians (modified from Wilson, 2013; Wilson et al., 2021); and 311 specimens (164 P3-4, 147 p3-4) from 18 species of late Neogene-Quaternary African, Asian and North American equids (modified from Barrón-Ortiz et al., 2017). For each dataset, we superimposed GM configurations using the generalized least-squares Procrustes (GLSP) method. We evaluated three proposed coding methods: PC component scores as continuous characters (PC-P), landmark coordinates as continuous characters (LC-P), and "Landmark Analysis Under Parsimony" (LAUP) via spatial optimization. All analyses assumed equal weighting, given the uncertain behavior of morphometric characters under extended implied weighting. No specimen-level analysis of either dataset recovered sampled species as reciprocally monophyletic. Trees exhibited incongruence with published morphological and/or molecular topologies, especially those produced using LC-P. Most nodes had low (if any) bootstrap, jackknife, or symmetric resampling support, especially on trees produced using PC-P. These observations are unsurprising given the extremely low number of configurations sampled in each dataset, as well as general issues with cladistic analyses of GLSP superimposed configurations. Still, analyses of the Cretaceous-Paleogene therian dataset provided more promising results compared to the equid dataset. We suspect tribosphenic dentition from "primitive" therians is more appropriate for specimen-level analyses of GM data than hypsosenodont or plagiolophodont dentition from ungulates, due to less extreme enamel wear through ontogeny. We recommend future cladistic analyses of GM data sample 3D rather than 2D configurations, to maximize shape information without oversampling vertices; consider other superimposition methods (e.g., resistant-fit theta-rho, dynamic alignment); investigate the effects of extended implied weighting; and overcome computational limitations of LAUPs.

SPECHT, CHELSEA¹ AND PATRICK O'GRADY²

¹LH Bailey Hortorium, Plant Biology Section, SIPS, Cornell University, USA and ²Department of Entomology, Cornell University, USA. -Corresponding author: cdspecht@cornell.edu

Vicki Funk: The Complete Botanist

We will address the many contributions to phylogenetics, neotropical botany and systematics of the members of the angiosperm family Compositae of the versatile and respected botanist Vicki Funk (1947-2019)

SRIVATHSAN, AMRITA¹, VIVIAN FENG¹, EMILY HARTOP¹, LESHON LEE², AND RUDOLF MEIER^{1,2}

¹Center for Integrative Biodiversity Discovery, Leibniz Institute for Evolution and Biodiversity Science, Museum für Naturkunde Berlin; ²Department of Biological Sciences, National University of Singapore. -Corresponding author: asrivathsan@gmail.com

Rapid, large-scale species discovery and descriptions using a reverse workflow through MinION barcoding with user-friendly software

The field of systematics faces significant challenges, given that over 80% of the animal species are unknown, undescribed and unidentifiable in middle of a biodiversity crisis. New and innovative approaches are needed that allow for the discovery and description of many unknown species. We use high-throughput integrative taxonomy to change this. We utilize DNA barcodes as the first data source to expedite the process of species-level sorting using a "reverse workflow." This method involves sequencing all specimens collected from bulk samples obtained with standard sampling techniques, like Malaise traps, and then grouping them into putative species units for expert taxonomists to examine. To facilitate this process, we must obtain barcodes from thousands of specimens—a task made feasible by the MinION, a real-time sequencing device from Oxford Nanopore Technologies. Portable and affordable, with an initial setup cost of less than \$5,000, the MinION allows for decentralized barcoding on a global scale when paired with low-cost and simple molecular methods for specimen processing. In this talk, I will present a variety of user-friendly software tools developed for large scale species-discovery using MinION, as well as tools that enable the reverse workflow. The focal point of the presentation will be ONTbarcoder2, a software featuring a graphical user interface that processes raw data from the MinION sequencer. This software generates consensus DNA barcodes, accounts for errors in the data, and performs quality control. By capitalizing on the real time sequencing, ONTbarcoder has recently been updated for real-time barcoding, allowing for the rapid identification of unknown specimens as sequences are generated. Furthermore, we will discuss recent improvements that enable us to barcode approximately 10,000 specimens per MinION flow cell at a cost of less than \$0.1 per barcode, as well as ~300 specimens from a smaller "Flongle" flow cell, costing roughly \$0.5 per specimen.

STEVENSON, DENNIS

New York Botanical Garden, NY, USA. -Email: dws@nybg.org

A Short History of Cladistics in Botany

This a brief foray into the history and persons involved in establishing cladistics in botany. It is neither comprehensive nor exhaustive. It covers roughly the period from 1980-1997. I will focus on early cladists and the papers, books, and symposia that had an influence in establishing cladistic approaches in botany and the conversion of the previous generation of some leading botanical systematists. It tracks doing analyses by hand to the development of programs that allowed for larger and larger matrices and the incorporation of new types of data such as molecular sequence data.

ST LAURENT, RYAN¹ AND PAUL Z. GOLDSTEIN²

¹Department of Entomology, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA; ²Systematic Entomology Laboratory, USDA, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA. -Corresponding author: ryanstlaurent93@gmail.com

*Phylogenomics challenge the current classification of Prominent Moths
(Lepidoptera, Notodontidae)*

The first phylogenomic results based on analyses of anchored hybrid enrichment (AHE) data from densely sampled tribes and subfamilies of Notodontidae or Prominent Moths will be presented. Using multiple phylogenetic inference methods (maximum likelihood, parsimony, and multi-species coalescent), we explore the backbone of the Notodontidae tree and its implications for the subfamily classification and our interpretation of key noctuid morphological features. We recognize 21 subfamilies of notodontids, eight for the first time as such. Our phylogenomic results (600+ loci) resolve several longstanding questions raised in Miller's (1991) morphological analysis, but contradict certain aspects of classifications based on that analysis and on the limited molecular analysis of Kobayashi and Nonaka (2016). Most surprisingly, our analyses reveal that Notodontidae is polyphyletic with respect to *Scrancia* Holland and relatives, a clade that has been variously recognized at the tribal and subfamilial rank within the Notodontidae. This group, currently recognized as *Scranciinae*, falls outside Notodontidae and appears to be the sister to the six established noctuid families. We will therefore elevate *Scranciinae* to family rank. Despite the identification of a novel lineage of Noctuoidea, deep-level relationships within the superfamily remain poorly established. The relative placement of the trifid families (Notodontidae, Oenosandridae, and the new family) with respect to the quadrifid families remains unclear. Following multiple analyses of several dataset configurations designed around the distribution of missing data, and an examination of strict support measures at the deepest nodes of the noctuid tree, we provisionally conclude that this ambiguity is a function of character conflict amplified by short branch lengths, i.e. rapid radiation.

SZUMIK, CLAUDIA AND MARÍA LAURA JUÁREZ

Unidad Ejecutora Lillo, Consejo Nacional de Investigaciones Científicas y Técnicas – Miguel Lillo 205, Tucumán, Argentina. -Corresponding author: szu.claudia@gmail.com

Thirty years of embiid phylogeny: stability in the light of inapplicable characters

Since the early 20th century, many studies on the behavior of Embioptera have shown the striking way of life of these insects; webspinners are gregarious and live their entire life inside silk tubes that they construct. This relictual Order contains approximately 500 described species from around the world, but a diversity of more than 1500 species is estimated. In the 20th century in terms of the taxonomy and systematics of the webspinners we have had the strong influence of two entomologists C. Davis (1913-1944) and E. Ross (1915-2016). From the 90's on, with the advances on phylogenetic methodology, several cladistic analyses were published; those studies generally focused on particular taxonomic levels (genera, family, order) and/or on including new evidence. Thanks to these quantitative analyses, a general cladogram of the order is becoming more congruent and stable. In addition, the different groups obtained in the last phylogenetic study, with the largest number of characters, are supported not only by male genitalia characters, but also by characters in other structures such as head, wings, female terminalia, and leg chaetotaxy. We propose here a review of these studies, using new advances in methods for analysis of inapplicable characters, all the known embiid genera as well as over 80% of its species, and a host of new characters.

TORRES, AMBROSIO¹, PABLO GOLOBOFF^{2,3}, AND SANTIAGO CATALANO^{2,4}

¹Center for Integrative Biodiversity Discovery (Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science), Berlin, 10115 Germany; ²Unidad Ejecutora Lillo, Consejo Nacional de Investigaciones Científicas y Técnicas - Fundación Miguel Lillo, Miguel Lillo 251, San Miguel de Tucumán, Tucumán, 4000 Argentina; ³American Museum of Natural History, 200 Central Park West, New York, NY, 10024 USA; ⁴Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán, Miguel Lillo 205, San Miguel de Tucumán, Tucumán, 4000 Argentina.- Corresponding author: atorresgalvis@gmail.com

Parsimony analysis in phylogenomics

Maximum Likelihood (ML) and Bayesian Inference (BI) are widely used concatenation-based methods in phylogenomics. This is possibly due to the offer of explicit models of molecular evolution, allowing for probabilistic hypothesis testing and statistical inference of parameters beyond topology (e.g., divergence times, rates of substitution). In contrast, Maximum Parsimony (MP) sees less frequent use, despite its faster computational speed and favorable performance under certain phylogenomic conditions (e.g., high levels of missing data, heterotachy, and Incomplete Lineage Sorting). However, it is worth noting that the choice among analytical methods often relies more on theoretical preconceptions rather than an extensive evaluation of the impact of methodological decisions on the resulting hypotheses. Here we summarize the key findings from a comprehensive comparative meta-analysis of ML, BI, and MP using 157 large empirical phylogenomic datasets, discussing the role of parsimony in modern molecular phylogenetics. The resulting topologies from ML, BI, and MP were very similar, sharing 89.1%—96.7% of the nodes. The differences were strongly associated with low-supported nodes, and proportionally larger in analyses at higher taxonomic levels (i.e., phyla and kingdoms). Factors such as high levels of missing data and long branches contribute to incongruence among methods, while the Guanine-Cytosine (GC) content showed no correlation with the incongruence. Importantly, we found a significant association between gene-tree incongruence and differences among optimality criteria, highlighting the potential of using gene-tree incongruence metrics in conjunction with site resampling analysis to predict incongruence among methods. We found that implied weighted parsimony analysis with appropriate weight settings yielded more congruent results with ML, whereas non-uniform transformation costs produced opposite outcomes. We emphasize the incongruence between morphological and molecular phylogenies and the limited availability of morphological phylogenies for comprehensive comparative analyses. Concluding, the prevailing inclination towards a specific method in phylogenomics lacks empirical support, even more so if we consider that the systematic biases assessed here, which contribute to method incongruence, can influence both MP and model-based approaches. Hence, we strongly recommend employing multiple criteria for method selection in phylogenomics, considering computational efficiency as a primary determinant. Additionally, conducting post-phylogenetic inference assessments becomes crucial to identify potential incongruent nodes among methods and evaluate the reliability of nodes and inferred phylogenetic relationships. Finally, we introduce a series of scripts designed for analyzing genomic datasets in TNT (Tree Analysis using New Technology). These scripts facilitate data concatenation, generation of graphs and datasets with different gene/taxon occupancy levels, support calculation, and phylogenetic reconstruction using concatenated or

individual gene matrices. They prove especially valuable for newcomers to TNT and enable the analysis of various data types beyond molecular alignments.

TORRES, AMBROSIO¹, AMRITA SRIVATHSAN¹ AND RUDOLF MEIER^{1,2}

¹ Center for Integrative Biodiversity Discovery (Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science), Berlin, 10115 Germany; ² Department of Biological Sciences, National University of Singapore, Singapore, Singapore. -Corresponding author: atorresgalvis@gmail.com

Molecular diagnoses for high-throughput taxonomy

New species descriptions require diagnoses. This means for species described with integrative methods (“integrative taxonomy”) should present diagnoses for all the different data types. However, many such species descriptions only present the barcode for the type or consensus sequences for the species; i.e., the descriptions lack molecular diagnoses. Here, we propose a new method for deriving and testing Molecular Diagnostic Combinations (MDCs). This method assigns a weight to each site based on the Jaccard index (i.e., the number of sequences in which the site states differ from the site state in the query taxon). Afterward, it uses a weighted random sampling approach to build the candidate combinations to become MDCs. At the same time, the method uses a stability measure of the specificity of the candidate combinations to assess their reliability. This measure relies on the suboptimal match values of each one of the combinations with the sequences that do not belong to the query taxon (the user can select the maximum suboptimal match accepted). As a final step, with the most frequent sites in the combinations that meet the preceding criterion, the final MDC is built (which must also meet the same criterion). We implemented the previously described method using the R programming language (script version and Shiny app version available at <https://atorresgalvis.shinyapps.io/MolecularDiagnoses/>). In addition, we included a module for testing MDC by applying it to all known sequences of a taxon. Using different empirical datasets, we demonstrate that the proposed method is efficient for obtaining reliable MDCs, which in turn are helpful in identifying unknown specimens. Finally, we briefly discuss some theoretical and philosophical concerns raised by some traditional taxonomists regarding the use of composite molecular characters for diagnoses and taxonomic descriptions.

WILL, KIPLING¹; DAVID MADDISON², WENDY MOORE³, KOJUN KANDA⁴, ALEXANDER L. WILD⁵, ANTONIO GOMEZ⁶, JAMES M. PFLUG², OLIVIA BOYD², KELLY MILLER⁷, AND JAOCHIM SCHMIDT⁸

¹University of California, Berkeley, USA; ²Oregon State University, USA; ³University of Arizona, USA; ⁴SEL, USDA, Smithsonian Institution, USA; ⁵The University of Texas at Austin, USA; ⁶Syracuse University, USA; ⁷The University of New Mexico, USA; ⁸University of Rostock, Germany. -Corresponding author: kipwill@berkeley.edu

A fossil-calibrated phylogeny of the beetle suborder Adephaga with an emphasis on Carabidae (Coleoptera)

A multigene phylogenetic analysis of the beetle suborder Adephaga based on an exemplar sample of 658 species is time calibrated using 41 selected fossils. Around 700 putative adephagan fossils are known but most are unsuitable for use in calibration. The criteria used for fossil inclusion and state of knowledge of carabid beetle fossils is presented. We compare the results, underlying assumptions, and limitations of several Bayesian approaches using a restricted set of the highest quality fossils and the full set of fossils that includes several apparently important but controversial specimens. Some underlying assumptions of the Fossil Birth-Death model developed to work well for small, well understood groups such as mammals, are found to be problematic for empirical datasets like ours, which is a relative small sample of a highly diverse clade. How this impacts the already problematic issue of maximum clade age is discussed. How diversification of defensive chemistry represented in the time-calibrated carabid phylogeny correlates with patterns of environmental change and potential ecological interactions in the past is presented.

WHEELER, WARD C.

Division of Invertebrate Zoology American Museum of Natural History, New York, NY, USA.
Email: wheeler@ammh.org

Multi-Armed Bandits, Thompson Sampling, and Machine Learning in Phylogenetic Graph Search

The multi-armed bandit problem is applied to phylogenetic graph searching. Thompson sampling is applied to a collection of “search-bandits” to favor productive search strategies over those that are less successful. This adaptive random sampling strategy is shown to be more effective in producing heuristically optimal phylogenetic graphs and more time efficient than existing randomized search strategies. The strategy is effectively a form of unsupervised machine learning that can be applied to a diversity of phylogenetic data sets without prior knowledge of their properties.

WILKINSON, MARK

The Natural History Museum, London, UK. -Email: m.wilkinson@nhm.ac.uk

Concatabominations

Rogue taxa can be phylogenetically unstable because of limited and extensive missing data. Their inclusion in phylogenetic reconstructions often leads to multiple trees, unresolved consensus topologies and an increase in run times. Safe Taxonomic Reduction (STR) has been used as an a priori method to determine the taxa that are potentially unstable and safe to exclude (in the sense that their exclusion can have no impact upon relationships inferred among the remaining taxa) from an analysis, however it has limitations and does not always solve unstable taxon problems. “Concatabominations” is a heuristic extension to STR based on the experimental combination of data from pairs of “potential taxonomic equivalents” to create new “concatabominated” taxa. Conceptually each concatabomination is analogous to forcing taxa

together in a tree. Compatibility is then used to investigate whether the concatabomination adds any homoplasy to the original data. Taxa that can be combined with many others without introducing any additional homoplasy to the data are candidate rogue taxa. A pipeline implementing the approach allows visualisation of taxonomic equivalence relations as connections between taxa in a network. Using gap-rich paleontological and gene tree matrix representation, concatabominations can outperform STR and increases the resolution of the resulting phylogenies. The approach is potentially significant for identifying and thus for addressing cases of ineffective taxonomic overlap in phylogenomic datasets. Use of parsimony scores rather than compatibility may further improve this approach.