3rd Annual Meeting of SPP 1991: Taxon-Omics

28 – 29 June 2019 in Munich

Program

We/Th, 26/27 June: Workshop "Hybridization capture, hyRAD, and targeted capture: workflows in the wet and dry lab"

Friday, 28 June

09:00 – **Morning coffee (catered)**; setting-up posters; they will be displayed throughout the meeting

10:00 – 11:00 Meetings of lab groups, early-career participants, and potential new participants

11:00 – 11:15 **Philipp Hühn, Markus S. Dillenberger, Michael Gerschwitz-Eidt, Jessica Los, Thibaud Messerschmid, Claudia Pätzold, and Gudrun Kadereit**, Mainz and Goettingen: Double digest RADSeq for long loci in phylogenomics – potential applications

11:15 – 11:30 **Nicolas Straube, Mariana Lyra, Axel Barlow, Michaela Preick and Michael Hofreiter**, Munich: Sequencing archival DNA from vertebrate wet collections

11:.30 – 11:45 **Janna Peters, Sven Rossel, Silke Laakmann, Pedro Martínez Arbizu, and Jasmin Renz**, Hamburg: Assessing zooplankton biodiversity by proteomic fingerprinting

11:45 – 12:00 **Michael Kloster, Bánk Beszteri and Tim W. Nattkemper**, Bielefeld: Combining virtual slides, web-based (multi-)expert annotation and image analysis for assembling training image sets for digital diatom analysis

12:00 – 13:00 **Lunch (catered in our building) and poster session**

13:00 – ... Summary of activities & resume of the past three years (**Renner and other steering committee members**)

14:00 – 15:00 Professor Wolfgang Wägele, Director of the Museum König in Bonn: Biodiversity monitoring requires digital species data and automated identification tools

15:00 – 15:30 **Coffee break (catered)**

15:30 – 15:45 **Bernhard Hausdorf and Christian Hennig**, Hamburg: Species delimitation and geography

15:45 – 16:00 **Torsten Hauffe, Jens Schauer, and Thomas Wilke**, Gießen: Towards a probabilistic and automated species-discovery system

16:00 – 16:15 **Karbstein, K., Tomasello, S., Hodac, L., Daubert, M., Pätzold, C. and Hörandl,** E., Göttingen: What is a species? Unraveling biodiversity within the agamic *Ranunculus auricomus* polyploid complex using RADseq and target enrichment

16:15 – 16:30 **Franziska Patzold, Anna K. Hundsdoerfer**, and six others, Dresden: Initial procedure in phenomics of Lepidoptera museum specimens to compare wing pattern genes using NGS

16:30 – 16:45 **Sabine Schiwitza and Frank Nitsche**, Cologne: Integrative taxonomy of protists exemplified by choanoflagellates

16:45 – 17:00 **Dirk Albach and Jannes Höpke**, Oldenburg: Genotyping polyploids with RAD methods - Insights and challenges

17:00 – 17:30 **Break (no catering)**

17:30 – 18:00 Miguel Vences and Ivaylo Kostadinov: Introduction to GFBio as background to the roundtable on Saturday morning

18:00 ... **Dinner (catered in our building)**

Saturday, 29 June

08:30 – ... **Coffee (catered in our building)**

09:00 – 10:00 Roundtable on Data deposition efforts in our SPP. Participants: Miguel Vences, Dominik Begerow, Ivaylo Kostadinov, Janine Felden, and Tanja Weibulat

10:00 – 10:30 **Break (not catered)**

10:30 – 10:40 **Robin-Tobias Jauss, Susanne Walden, Michael Bonkowski, Martin Schlegel,** Leipzig: Deep molecular characterization of protist diversity

10:40 – 10:50 **Susanne Walden, Martin Schlegel, Michael Bonkowski**, Cologne: Primer design and trouble-shooting: When general eukaryotic primers don't work

10:50 – 11:00 **Loïs Rancilhac, Iker Irisarri, Hervé Philippe, and Miguel Vences**, Braunschweig: A history of reticulation and inter-lineage gene flow, reconstructed by a phylotranscriptomic analysis of salamanders and newts (Salamandridae)

11:00 – 11:10 **Lena Steins and Dominik Begerow**, Bochum: Comparative genomics of smut fungi indicate incapability of meiotic division in life cycle reduced Ustilaginomycetes

11:10 – 11:20 **Fabian Deister, Sergio Vargas, Dirk Erpenbeck, and Gert Wörheide**, Munich: Comparative genomics and the nature of species in sponges - case studies in genus *Tethya*

11:20 – 11:30 **Darrin Hulsey and Axel Meyer**, Konstanz: The genomic basis of diagnostic species phenotypes: will phylogenomics help (or hurt) our taxonomic inferences?

11:30 – 11:40 **Agnes Scheunert, Ulrich Lautenschlager, Florian Wagner, Tankred Ott, Marco Dorfner, Robert Vogt, and Christoph Oberprieler**, Regensburg: Approaches for tracking polyploidization and hybridization in the young, closely related *Leucanthemum* complex (Compositae, Anthemideae)

11:40-12:00 OPEN discussion & official farewell

12:00-13:00 Lunch (catered in our building)

Genotyping polyploids with RAD methods - Insights and challenges

Dirk C. Albach and Jannes Höpke

Institute for Biology and Environmental Sciences, Carl von Ossietzky-University Oldenburg; email: dirk.albach@uol.de

Genomic analyses of polyploids face a number of complications compared to analyses of diploids. Except for a few model organisms, the distinction between orthologous and paralogous loci is almost impossible. Thus, the presence of up to four alleles per locus is possible. Most software designed for SNP analysis neglects this and transforms data to pseudo-diploid data. For interpretation of results from our analysis of Genotyping-by-Sequencing (GBS)-data in *Veronica* subg. *Pseudolysimachium*, we simulated data for mixed ploidy-populations under different migration rates and with different coding of heterozygosity. Our results demonstrate little effect of diploidization of data but question results based on dominant markers. We also investigated the effect of paralogs as identified using paralog-finder software. Fortunately, the inclusion of paralogs does not seem to introduce a bias. However, these are just some factors influencing the results of GBS-analyses with mixed ploidy-data and require further careful analyses of GBS- or similar data.

Comparative genomics and the nature of species in sponges - case studies in genus *Tethya* (Porifera)

Fabian Deister[1], Sergio Vargas[1], Dirk Erpenbeck[1, 2], Gert Wörheide[1, 2, 3]

[1]Department of Earth and Environmental Sciences, Palaeontology & Geobiology, Ludwigs-Maximilians-Universität München.
[2]GeoBio-Center, Ludwigs-Maximilians-Universität München
[3]SNSB–Bayerische Staatssammlung für Paläontologie & Geologie

In sponges, species-identification and species delimitation are often difficult. Most frequently, a morpho-species concept is used to diagnose species but since diagnostic morphological characters are often rare or variable even within species, it remains unclear what a species actually is in sponges. While some animal species can be precisely identified via commonly used barcode genes, barcoding sponges remains challenging. Since morphology and classic DNA barcoding have reached their limit and often cannot be used to diagnose sponge species, we aim to go further and use comparative genomics to diagnose and identify sponge species. Next generation sequencing allows for the *de novo* sequencing of animal genomes, which enables the discovery of potentially suitable species-delimiting markers. As a test case, we sequenced the draft genomes of three species in the genus *Tethya* using Illumina and Oxford Nanopore sequencing technologies. We use the assembled genomes and supporting RNA data to identify, annotate and compare possible species-specific marker loci. To support the identification of orthologous loci, we searched for collinear, syntenic protein-coding regions and identified probable species specific rearrangements and inversions between two of the species sequenced (*Tethya wilhelma* and *Tethya minuta)*. Here we present a methodology developed to search for syntenic clusters and rearrangements between genomes. We show results for ~10,000 single copy orthologs found in those two species and discuss whether genome structure differences can be used to distinguish between these two congeneric species.

Microgastropod Taxon-Omics: Towards a probabilistic and automated species-discovery system

Torsten Hauffe[1], Jens Schauer[1,2], Thomas Wilke[1]
[1]Systematic & Biodiversity Group, Justus Liebig University Giessen, Germany, [2]Bioinformatics & Systems Biology, Justus Liebig University Giessen, Germany; email: torsten.hauffe@bio.uni-giessen.de

Significantly reducing the share of undocumented species is not trivial because there is an inherent conflict between the two main interests of taxonomy – quality of delimitation/description and speed of delimitation/description. We address both aspects by developing a probabilistic species-discovery system (proSDS) that groups specimens into species, informs about the associated probability, and enables the

detection of unknown species ('novelty detection'). This is accomplished by applying classification rules derived from supervised machine learning on a reference database of genetic, morphological, and ecological traits. Our proSDS uses artificial neural networks and support vector machines for classification. To evaluate the classification performance, we simulated sequences and morphological/ecological traits along species trees and added population variability, resulting in various ratios of intra/interspecific variation. Our initial results showed no performance difference between both machine learning algorithm but a decrease in classification precision with increasing intraspecific genetic variation. Unfortunately, this decreasing performance could not be compensated for by morphological or ecological traits. We therefore implemented a 'feature selection', which filters uninformative and/or noisy traits via cross-validation. This resulted in the stabilization of classification precision throughout a wide range of intraspecific genetic variation. With the gained confidence in the robustness of proSDS, we could successfully demonstrate a high classification precision for two genera of freshwater microgastropods. Our approach might assist scientists in making taxonomic decisions by estimating the probability for a specimen of unknow species identity to belong to a known or novel species.

Species Delimitation and Geography

Bernhard Hausdorf[1,*] and Christian Hennig[2]

[1]*Center of Natural History, University of Hamburg, Germany*; *email: hausdorf@zoologie.uni-hamburg.de*
[2]*Dipartimento di Scienze Statistiche "Paolo Fortunati", Università di Bologna, Italy*

Despite the importance of the geographical arrangement of populations for the inference of species boundaries, only few approaches that integrate spatial information into species delimitation were developed so far. Persistent differentiation of sympatrical groups of individuals is the best criterion for species status. Species delimitation becomes much catchier if allopatric populations are considered because of geographical variation. It is often difficult to assess whether differences between allopatric populations might persist upon contact. We propose a novel method for testing the hypothesis that the genetic distances based on multilocus markers between individuals or populations belonging to two different predefined groups are not larger than expected based on their geographical distances and the relationship of genetic and geographical distances within the predetermined groups. A rejection of this null hypothesis provides evidence for the species status of the two studied groups. Case studies show that the proposed tests based on distances between individuals are suitable to distinguish between intra- and interspecific differentiation. The regression approach proposed here is more appropriate for testing species hypotheses with regard to isolation by distance than (partial) Mantel tests.

The genomic basis of diagnostic species phenotypes: will phylogenomics help (or hurt) our taxonomic inferences?

Darrin Hulsey and Axel Meyer

Department of Biology, University of Konstanz; email: darrin.hulsey@uni-konstanz.de

Using phylogenomics to inform taxonomy might be less straightforward than has been previously thought. Genomic studies of the traits used both to diagnose species and that might often promote reproductive isolation have highlighted that a few small gene regions could generate the differences that both organisms themselves as well as humans often use to diagnose taxa. The remainder of the genome, constituting effectively tens of thousands of loci, might rarely mirror the evolution of these loci underlying reproductive isolation and taxonomic diagnoses. I will highlight several misleading examples and briefly open discussion to how we should best move forward if we want to marry genomics and taxonomy.

Deep Molecular Characterisation of Protist Diversity in Forest Soils and Tree Canopies

Robin-Tobias Jauss[1], Susanne Walden[2], Michael Bonkowski[2], Martin Schlegel[1,3]

[1]Molecular Evolution & Animal Systematics, University of Leipzig, Talstraße 33, 04103 Leipzig
[2]Institute of Zoology, Department of Terrestrial Ecology, University of Cologne, Zülpicher Straße 47b, 50674 Köln
[3]German Centre for Integrative Biodiversity Research (iDiv) Halle Jena Leipzig, Deutscher Platz 5e, 04103 Leipzig, jauss@uni-leipzig.de

Protists occupy key nodes in terrestrial food webs due to their high abundance, fast turnover and functional importance as microbial grazers. However, methodological drawbacks in both culturing and molecular methods still strongly limit the knowledge of protist diversity, so that large groups remain virtually unknown. Here, we apply advanced cultivation-independent high throughput sequencing methods using newly designed group-specific primers and reverse rRNA-transcriptomics for a comprehensive assessment of protist diversity across all ecological compartments from forest soils (litter layer & mineral soil) to the canopy region (bark, leaves, dead wood, branch forks, knotholes, epiphytes) in temperate and tropical biomes. Our analyses yielded more than 600 oomycete OTUs. Diversity patterns show a distinct difference in community composition depending on the strata ground and canopy, indicating that stratification shapes oomycete diversity to a greater extent than microhabitat-filtering. This, however, seems to be different for the Cercozoa. Here, unique microhabitat specific communities reveal filtering effects which now have to be put into a taxonomic and ecological context.

'What is a species'? Unraveling biodiversity within the agamic *Ranunculus auricomus* polyploid complex using RADseq and target enrichment

Karbstein, Kevin[1,2,*], Tomasello, Salvatore[1], Hodac, Ladislav[1], Daubert, Mareike[1], Pätzold, Claudia[1] & Hörandl, Elvira[1]

[1] Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Goettingen, Germany
[2] Georg-August University School of Science (GAUSS), Goettingen, Germany
[*] email: kevin.karbstein@uni-goettingen.de

Recognizing and delimiting species is methodically challenging in fast evolving, evolutionary young species complexes characterized by recent, steady gene flow/introgression and incomplete lineage sorting (ILS). We propose an innovative workflow by combining analysis based on optimized RADseq, target enrichment and geometric morphometric data to disentangle such species groups. The *Ranunculus auricomus* complex is one of the largest apomictic polyploid complexes (> 800 species) in Europe but taxonomically widely unresolved. As proposed by Hörandl (2018) for agamic species complexes, we first analyzed the sexual species (see Dunkel et al. 2018) of *R. auricomus*. Concatenated RAxML/ExaML tree based on RADseq data, coalescent species tree based on target enrichment (TE), and STRUCTURE analysis revealed two well-resolved clades, i.e. a basal clade with non-dissected leaf species ('*cassubicus*') and a derived clade with non-dissected ('*cassubicus*') and dissected ('*auricomus*') leaf species rejecting the old Linnaean complex classification into the two species ('*auricomus*', '*cassubicus*'). Thus, non-dissected leaf lineages have been the progenitor of all, morphologically diverse, dissected leaf lineages. Furthermore, we used the longer, single-copy regions produced by TE to explore temporal patterns of species diversification by producing time calibrated phylogenies. We also examined RaXML/ExaML tree by quartet sampling method (Pease et al. 2018) to pinpoint tree discordances and possible causes of low node support (gene flow/introgression, ILS). Finally, we intend implementing sequence-based species delimitation methods (e.g., SNAPP) and geometric morphometric analysis to disentangle the sexual species and propose a well-founded species concept.

References

Dunkel F.G., Gregor T., Paule J. 2018. New diploid species in the Ranunculus auricomus complex (Ranunculaceae) from W and SE Europe. Willdenowia. 48:227–257.

Hörandl E. 2018. The classification of asexual organisms: Old myths, new facts, and a novel pluralistic approach. Taxon. 67:1066–1081.
Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am. J. Bot. 105:385–403.

Combining virtual slides, web-based (multi-)expert annotation and image analysis for assembling training image sets for digital diatom analysis

Michael Kloster[1], Bánk Beszteri[2] and Tim W. Nattkemper[3]

[1]Faculty of Technology, University of Bielefeld; email: michael.kloster@uni-bielefeld.de
[2]Faculty of Biology, University Duisburg-Essen; email: bank.beszteri@uni-due.de
[3]Biodata Mining group, Faculty of Technology, University of Bielefeld; email: tim.nattkemper@uni-bielefeld.de

Computational taxonomic identification of specimens in light-microscopic images is a challenging application field for modern deep learning algorithms. Deep network training requires a large number of training samples, which can be a limiting factor as data collection is labor-intensive and requires taxonomic expert knowledge. To overcome this bottleneck of training data collection we combine slide-scanning microscopy, collaborative virtual slide annotation and semi-automated object detection into an integrated workflow enabling us to generate large numbers of high-quality annotated training data sets on a new level of efficiency. This workflow comprises selected ImageJ software components for virtual slide stitching, BIIGLE (www.biigle.de) for web-based annotation, SHERPA (www.awi.de/sherpa) for object detection and the newly developed tool SHERPA2BIIGLE for connecting the latter two and handling all functions needed for generating training data sets. The applicability of this workflow has been successfully tested on diatom material originating from Southern Ocean net samples.

GFBio: FAIR Data Services for Biodiversity, Ecology and Environmental Science

Ivaylo Kostadinov, GFBio Consortium

GFBio - Gesellschaft für Biologische Daten e.V.; email: ikostadi@gfbio.org

The German Federation for Biological Data (GFBio, www.gfbio.org) was established as a sustainable, national infrastructure to support researchers though the complete life cycle of research data management, from planning, through acquisition, to publication [1].

GFBio brings together experts from various fields, including natural history collections, environmental data publishing and bioinformatics. Our services to the community include:

Data Management Plan (DMP) Support - data management is an integral part of the research process. Consequently, funding agencies now require a DMP as part of grant applications. Together with our data management experts, you can prepare a custom DMP for your next proposal or ongoing project.

Data Search, Visualization and Analysis - GFBio offers an integrated metadata search across its associated data centers. Data can be downloaded collectively or visualized together with additional layers of your own choosing.

Data Annotation, Archiving and Publication - the GFBio data brokerage service offers a single point of contact for archiving biological data in one (or several) of our associated, long-term data centers. The added value for our users includes manual curation, annotation and standardization support.

1. Diepenbroek, et al. 2014. "Towards an Integrated Biodiversity and Ecological Research Data Management and Archiving Platform : The German Federation for the Curation of Biological Data (GFBio)." In Informatik 2014 – Big Data Komplexität Meistern. GI-Edition: Lecture Notes in Informatics (LNI) - Proceedings, edited by E Plödereder, L Grunske, E Schneider, and D Ull, P-232:1711–24. Bonn: Köllen Verlag. http://subs.emis.de/LNI/Proceedings/Proceedings232/1711.pdf.

Initial procedure in phenomics of Lepidoptera museum specimens to compare wing pattern genes using 'Next Generation Sequencing' (NGS)

Franziska Patzold, Senckenberg Naturhistorische Sammlungen Dresden, Germany, Email: franziska.patzold@senckenberg.de, Tomasz Suchan, Institute of Botany, Polish Academy of Sciences, Poland, Email: t.suchan@botany.pl, Sylke Winkler, Max Planck Institute of Molecular Cell Biology and Genetics, Germany, Email: winkler@mpi-cbg.de, Wenhua Tan, Max Planck Institute of Molecular Cell Biology and Genetics, Germany, Email: tan@mpi-cbg.de, Gene Myers, Max Planck Institute of Molecular Cell Biology and Genetics, Germany, myers@mpi-cbg.de, Martin Pippel, Max Planck Institute of Molecular Cell Biology and Genetics, Germany, Email: pippel@mpi-cbg.de, Michael Hiller, Max Planck Institute of Molecular Cell Biology and Genetics, Germany, Email: hiller@mpi-cbg.de, **Anna K. Hundsdoerfer,** Senckenberg Naturhistorische Sammlungen Dresden, Germany, Email: anna.hundsdoerfer@senckenberg.de

Hybridization is frequent in the Palearctic *Hyles* species and intermediate phenotypes, such as transitional forewing patterns, occur regularly. To cover large parts of the species' distribution areas, we study museum exemplars. The superior DNA extraction method targets single stranded DNA. We used two different approaches of the hyRAD protocol (Hybridization Restriction-Associated-DNA sequencing) to obtain single nucleotide polymorphisms and compare genomic compositions. To determine the differences in endogenous DNA content, contamination and detected single nucleotide polymorphisms, we further compared these two protocols to shotgun sequencing. This comparison will help to further evaluate and utilize modern high throughput techniques for the use of museum samples in phylogenomic studies in Lepidoptera. We sequenced a high quality genome of the genus *Hyles* from fresh tissue using PacBio long-read sequencing technology. The contigs are aligned to the genome assembly of *Bombyx mori* to putatively assign them to chromosomes. Well known wing pattern genes like *optix*, *cortex* and *wntA* can then be searched at the expected sites and compared between *Hyles* species'. This provides a promising opportunity to document and understand changes in biodiversity, gaining insight into the evolutionary process of speciation.

Assessing zooplankton biodiversity by proteomic fingerprinting

Janna Peters, Sven Rossel, Silke Laakmann, Pedro Martínez Arbizu, and Jasmin Renz

German Center for Marine Biodiversity Research, Senckenberg Institute, Germany; email: janna.peters@senckenberg.de

Taxonomic knowledge is a key premise for reliable zooplankton monitoring. However, for many calanoid copepod species identification to species level remains challenging and too time-consuming for routine analysis. This is even more difficult in biodiversity assessments for regions with a high degree of undescribed new species. Species identification using specific mass profiles of peptides and small proteins (proteome fingerprints) is a time- and cost-efficient alternative to molecular discrimination techniques and well established in microbiology. This study aims to evaluate the applicability of this method for the identification of abundant copepod species in the ICES monitoring area. Using random forest algorithms, we tested for stability of species-specific signals between different copepod populations, including 30 species from 15 different regions. Less than 1% of individuals were mis-identified. No phylogenetic signal was detected in hierarchical cluster analysis. In a next step unsupervised k-medoids cluster analysis was tested to identify species composition and assess biodiversity without a reference data base and without prior taxonomic identification. The final method was then applied to a data set of bentho-pelagic deep sea copepods from the South Atlantic and compared with COI based species composition. Preliminary data suggest that due to species richness and a high degree of singletons in this habitat this method would lead to an underestimation of biodiversity. Further analyses for optimization of unsupervised classification are planned. In conclusion, proteomic fingerprinting is considered to be a promising tool to support studies on copepod biodiversity.

A history of reticulation and inter-lineage gene flow, reconstructed by a phylotranscriptomic analysis of salamanders and newts (Salamandridae)

Loïs Rancilhac, Iker Irisarri, Hervé Philippe, and Miguel Vences

Zoological Institute, TU Braunschweig; e-mail : loisrancilhac@gmail.com

Using both complete mitochondrial genomes and transcriptomic data, we investigated the phylogenetic relationships of the Salamandridae (Amphibia:Caudata), obtaining two strongly supported topologies. However, comparing them confirmed a series of discrepencies between both markers, already suggested by previous studies. A careful analysis of the genomic data, using both introgression tests and phylogenetic networks analyses, yielded strong signal for introgression between several lineages, suggesting that some of the aforementioned discordances resulted from mitochondrial introgression following gene flow events.

Approaches for tracking polyploidization and hybridization in the young, closely related *Leucanthemum* complex (Compositae, Anthemideae)

Agnes Scheunert, Ulrich Lautenschlager, Florian Wagner, Tankred Ott, Marco Dorfner, Robert Vogt, and Christoph Oberprieler

Evolutionary and Systematic Botany Group, Institute of Plant Sciences, University of Regensburg; email: christoph.oberprieler@ur.de

Hybridization and polyploidization are frequent phenomena in recently diverged species groups, which often renders their phylogenetic evaluation and the delineation of reliable taxonomic groups a challenging task. This is due to low amounts of phylogenetic signal, minor morphological differences and the impact of gene flow that blurs phylogenetic ancestry and relationships among polyploid and diploid members of a group. We assessed polyploidization and homoploid hybridization in the close-knit *Leucanthemum* complex on the molecular level using several approaches. One of those involved reconstruction of diploid ancestors of polyploids, through partition of their alleles into subsets corresponding to hypothetical ancestral taxa, in a way that minimizes deep coalescences (i.e., extra lineages required) in a resulting multi-labelled species tree. For the application of this approach to an enlarged dataset, 20 nuclear low-copy markers were designed which will be used as home-made probes for target capture and subsequent PacBio sequencing. Additionally, the first plastid genomes within Leucantheminae have been generated using Illumina and Oxford Nanopore sequencing and will serve as references for genome skimming of further *Leucanthemum* plastomes. Apart from targeting long, known nuclear or plastid markers, anonymous loci obtained from ddRAD sequencing have been de-novo assembled using ipyrad; NeighborNet networks were constructed from K2P distance matrices based on the concatenated SNPs. Results from both nuclear/plastid marker and RADseq data point towards an origin of all investigated polyploid taxa exclusively in one of two subclades of *Leucanthemum*; the analyses are indicative of allo- or autopolyploid speciation in some cases, however others remain inconclusive.

Integrative taxonomy of protists exemplified by choanoflagellates

Sabine Schiwitza and Frank Nitsche

Biological Department, Institute of Zoology, Biocenter Cologne, University of Cologne, Zuelpicher Straße 47b, 50674 Cologne, Germany, sabine.schiwitza@uni-koeln.de

Studies on protists have undergone a series of major paradigm shifts during the last century that have radically changed our concept of these organisms regarding their structural function in microbial food webs and our technological approaches how they can be studied. Latest innovative molecular techniques, as high-throughput sequencing, produce massive amounts of reads which expand vastly our knowledge on protist biodiversity patterns. But only few of these generated sequences can be assigned to any known eukaryotic group. Thus, an urgent need exists in terms of matching known morphospecies with anonymous sequence data in order to refine and augment the future knowledge output of the rapidly developing molecular toolbox. Within our study, we focus on an integrative taxonomical approach to extend common reference databases with encompassing species descriptions which is still fundamental regarding the interpretation of any metabarcoding study. These integrative data will allow for a comprehensive analysis regarding protist biodiversity, community structures and their ecological functionality. Comparative genomics of smut fungi indicate ability of meiotic division in the asexual genus Pseudozyma

Comparative genomics of smut fungi indicate incapability of meiotic division in life cycle reduced Ustilaginomycetes

Lena Steins and Dominik Begerow

AG Geobotanik, Evolution und Biodiversität der Pflanze, Ruhr-Universität Bochum; email: lena.steins@rub.de

Ustilaginomycetes are a group of host specific plant parasitic fungi that undergo an anamorphic yeast stage and a teleomorphic parasitic stage during their life cycle. However, species in the genus Pseudozyma do not have a described teleomorphic stage and are therefore viewed as asexual species. Using comparative genomics of 19 Ustilaginomycete species, we analyzed the occurrence and coupling of 27 core meiosis genes and mating loci to assess the question if these species have retained the genetic makeup for mating and meiosis. The presence of core meiosis genes in the genus Pseudozyma leads to the conclusion that these species have at least some ability to mate. Thus, this raises the question whether they have a parasitic stage or have recently switched to a non-parasitic life style as an evolutionary strategy, maybe after losing access to their host species.

Sequencing archival DNA from vertebrate wet collections

Nicolas Straube[1], Mariana Lyra[2], Axel Barlow[3], Michaela Preick[3] & Michael Hofreiter[3]

1 SNSB- Bavarian State Collection of Zoology, Ichthyology, Munich, Germany; email: **straube@snsb.de**
2 Laboratório de Herpetologia, Depto de Zoologia, Rio Claro, Brasil; email: **marillyra@gmail.com**
3 Evolutionary Adaptive Genomics, Institute for Biochemistry and Biology, Department of Mathematics and Natural Sciences, University of Potsdam, Germany; emails: **axel.barlow.ab@gmail.com**, **mpreick@uni-potsdam.de**, **michael.hofreiter@uni-potsdam.de**

Recent advances in next generation sequencing of ancient DNA are promising for application to museum wet collection material. Here we present results from experiments where we 1) applied ancient DNA extraction methods on different vertebrate wet collection tissues and 2) converted the extracted archival DNA (arcDNA) into single-stranded DNA libraries. We succeeded in extracting arcDNA from all tested samples and further accomplished to construct DNA libraries from those. Correlations of variables collected during laboratory work indicate an expected negative correlation between the age of samples and endogenous DNA content as well as average library size of the sequenced libraries. Several other significant correlations are useful when evaluating sample quality in the course of laboratory work, such as the detected negative correlation between the optimal number of PCR amplification cycles and average fragment size. As our focus was on formaldehyde fixed samples, we tested if heating the samples during DNA extraction could increase library complexity. Even though heating resulted in higher DNA concentration, library complexity did not increase. We also performed targeted gene capturing for the mitochondrial genome based on bait sequences generated from long range PCR products. Sequencing of target captured libraries allowed for reconstructing large parts of the mitochondrial genome of samples, where tRNAs and rRNAs are overrepresented. Subsequent phylogenetic analyses of included type specimens and material, which has never been sequenced before, already provide useful information for taxonomy.

Primer Design and Troubleshooting: Why do "General" Eukaryotic Primers Select Against Some Major Protist Taxa?

Susanne Walden (1), Martin Schlegel (2), Michael Bonkowski (1)

(1) University of Cologne, Institute of Zoology, Cologne, Germany; (2) University of Leipzig, Institute of Biology, Leipzig, Germany; E-Mail: s.walden@uni-koeln.de

Eukaryotic unicellular organisms (protists) play an important role in the food webs of aquatic and terrestrial environments and are a huge reservoir of unexplored biodiversity and functions. However, due to the tremendous diversity of the numerous deep phylogenetic lineages of protists, their communities cannot be recovered with a single universal barcode. The group of Amoebozoa, which has been found to be exceptionally abundant in terrestrial environments, is probably one of most understudied supergroups in the eukaryotic tree of life. Their high genetic divergence even within conserved barcoding regions lead to strong underrepresentation in both taxonomy and molecular surveys applying "universal primers". Consequently, this current state of research reveals only a fragmented and superficial view of their hidden diversity. Besides comparison between the results of PCR-free RNA sequencing and high-throughput DNA metabarcoding, application of group-specific primers is another reliable tool to avoid the drawbacks of "general" primers. To design highly group-specific primers that specifically amplify this major group of protists, we precisely analysed novel potential barcoding genes for Amoebozoa, in particular the mitochondrial cytochrome oxydase subunit 1 (COI) and the hypervariable region 4 (V4) within the SSU rDNA.