Leopoldina
Nationale Akademie
der Wissenschaften

Juni 2014 | Online-Supplement zur Stellungnahme

Herausforderungen und Chancen der integrativen Taxonomie für Forschung und Gesellschaft

Taxonomische Forschung im Zeitalter der *OMICS*-Technologien

Leopoldina
Nationale Akademie
der Wissenschaften

Herausforderungen und Chancen der integrativen Taxonomie für Forschung und Gesellschaft

Supplement zur Stellungnahme

Impressum

Herausforderungen und Chancen der integrativen Taxonomie für Forschung und Gesellschaft - Taxonomische Forschung im Zeitalter der OMICS-Forschung. Online-Supplement zur Stellungnahme.

Challenges and Opportunities of Integrative Taxonomy for Research and Society – Taxonomic Research in the Era of OMICS Technologies. Online Supplement to the Statement.

Herausgeber:
Deutsche Akademie der Naturforscher Leopoldina e.V.
– Nationale Akademie der Wissenschaften –
Geschäftsstelle: Jägerberg 1, 06108 Halle (Saale)
Berliner Büro: Reinhardtstraße 14, 10117 Berlin

Redaktion:
Prof. Dr. Rudolf Amann
Dr. Henning Steinicke

Zitiervorschlag:
Nationale Akademie der Wissenschaften Leopolina, 2014: Herausforderungen und Chancen der integrativen Taxonomiefür Forschung und Gesellschaft - Taxonomische Forschung im Zeitalter der OMICS-Forschung. Online-Supplement zur Stellungnahme. Deutsche Akademie der Naturforscher Leopoldina e.V. Nationale Akademie der Wissenschaften, Halle/Saale.

Vorbemerkung

Die Leopoldina-Arbeitsgruppe „Integrative Taxonomie im *OMICS*-Zeitalter“ hat im Rahmen ihrer Arbeiten insgesamt drei Workshops veranstaltet, bei der Arbeitsgruppenmitglieder und Gäste aus relevanten Forschungsbereichen angehört wurden. Am 21. und 22. Juni 2012 wurde in Halle/Saale ein Workshop mit dem thematischen Schwerpunkt „Taxonomie in der Botanik“ veranstaltet, gefolgt von einem Workshop am 10. und 11. September 2012 in Bremen zum Thema „Taxonomie in der Mikrobiologie, medizinischen Mikrobiologie und der Mykologie“ und einem letzten Workshop mit dem Titel „Taxonomie in der Zoologie“ am 13. und 14. Dezember 2012 in Berlin.

Das vorliegende Supplement zur Stellungnahme „Herausforderungen und Chancen der integrativen Taxonomie für Forschung und Gesellschaft – Taxonomische Forschung im Zeitalter der OMICS-Technologien“ beinhaltet die schriftlichen Beiträge der Referenten dieser Workshops, sofern sie der Akademie zur Verfügung gestellt wurden.

Die Akademie dankt allen Referenten und Autoren für ihre Beiträge.

Für den Inhalt der Textbeiträge und die korrekte Angabe der Quellen sind die Autoren verantwortlich. Die Beiträge sind teilweise in englischer, teilweise in deutscher Sprache verfasst.

Inhalt

Textbeiträge Workshop Botanik 6

Taxonomic challenges in biodiversity research – Helge Bruelheide 6

Taxonomy and –omics data: two sides of a coin or interdisciplinary linkage via databases, networks and technical solutions? – Birgit Gemeinholzer 13

Are the Nomenclatural Codes Fit for the Future? A Protistologist's View – Regine Jahn 17

TAXON-omics – Taxonomie in Deutschland im Zeitalter der „-omics-Forschung" – Volker Lohrmann 23

Main challenges in the taxonomy of land plants (embryophytes) – Susanne S. Renner 27

Phylogenomics versus phylophysiology – static versus dynamic – Wolfram Weckwerth 29

Textbeiträge Workshop Mikrobiologie, medizinische Mikrobiologie, Mykologie 32

Microbiomes and metagenomes in need of professional classification – Rudolf Amann 32

Fungal genomics and its impact on secondary metabolite discovery and elucidation of pathogenicity – Axel A. Brakhage 36

Fungal biodiversity with respect to carbon utilisation: Taxonomic relationships versus parallel evolution – Ronald P. de Vries 37

The polyphasic approach in the "omics" era – up to date, or out of date? – Peter Kämpfer 41

Entwicklung und Biotechnologie von Pilzen - Fungal development, fungal biotechnology – Ulrich Kück 44

Typing and identification of microorganisms at sub-species levels: considerations for biodiversity, biotechnological, clinical and epidemiological studies – Edward Moore 49

Diversification in nonpathogenic bacteria - implications for taxonomy, bioinformatics and data repositories – Jörg Overmann 50

Past, presence and future of microbial taxonomy – Ramon Rosselló-Móra 54

The intestinal microbiome and human health: a new challenge for taxonomy of metagenomic data – Bärbel Stecher 58

Characterizing species boundaries and species histories in closely related fungi using comparative population genomic approaches – Eva H. Stukenbrock 63

(Gen-)Omics and bacterial pathogens: Impact and Potential – Sebastian Suerbaum 67

Textbeiträge Zoologie-Workshop 71

Wolbachia and molecular barcoding - conceptions and misconceptions – Christoph Bleidorn 71

Ancient DNA – retrospective sequencing of types – Michael Hofreiter, Chenhong Li, Gavin Naylor, Tanja Haus and Kirsten Richter 75

Next generation DNA sequencing methods now permit the genetic delineation of extremely young (cichlid fish) species and populations – Axel Meyer 76

1KITE – 1K Insect Transcriptome Evolution – Bernhard Misof 77

The use of DNA sequence data in animal identification and classification – Michael J. Raupach 79

Natural hybridization in primates and what "-omics" contributed to primate taxonomy and systematics – Christian Roos, Dietmar Zinner 85

Barcoding microbial eukaryotes – Thorsten Stoeck 90

Integrative taxonomy: perspectives for simultaneously improving quality and speed of a global species inventory – Miguel Vences 95

The German Barcode of Life Project and the contribution of DNA barcoding to modernize taxonomy – J. Wolfgang Wägele 99

"The Age of Taxonomy" – Quentin Wheeler 102

Textbeiträge Workshop Botanik

Taxonomic challenges in biodiversity research

– Helge Bruelheide

Martin-Luther-University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden, Halle (Saale), Germany

iDiv - German Centre for Integrative Biodiversity Research, Leipzig, Germany

Introduction

In biodiversity (BD) research, taxon identification is an inherent element in all types of research, as virtually all types of BD data are taxon-based. These data range from databases on herbarium specimens (e.g. CVH 2012), zoological collections (e.g. Senckenberg 2012), species occurrences (e.g. GBIF 2012), and species geographical distributions (e.g. Chorology Database Halle, CDH 2012), species co-occurrence information (e.g. vegetation databases such as GVRD 2012, Schaminée et al. 2009) or all-taxa plot inventories (such as in the BD Exploratories, Fischer et al. 2010). In molecular BD screening surveys, DNA sequences are first processed and then clustered into operational taxonomic units (OTUs). In most of these studies, then an attempt is made to assign sequences to taxonomic databases (such as NCBI 2012, or specialized ones, e.g. for arbuscular mycorrhizal fungi, Öpik et al. 2010). The rationale to use taxa as the fundamental unit in BD research is the quest to anchor all observations made to something defined and permanent, thus empowering other researchers at different points in space and time to make use of these observations. This quest is particularly obvious when molecular gene genealogies differ between loci used (see contribution of Eva H. Stukenbrock in this volume), and genome-wide coalescence approaches are used to come to a consensus.

Anchoring BD observations in taxonomy is necessary for all questions asked in contemporary BD research: (i) how can we detect and quantify biodiversity? (ii) how does biodiversity emerge? (iii) what are the consequences of biodiversity for the functioning of ecosystems? (iv) how can we safeguard biodiversity (iDiv 2012). These questions go far beyond documentation of where organisms have been encountered (L matrix type of information, Fig. 1), but require links to other types of information (Fig. 1). On the one hand, taxa are linked to characteristics observed on them, commonly referred to as species traits (Q matrix in Fig. 1), on the other hand, the sites where taxa have been found will be interpreted ecologically, linking them to environmental data (R matrix in Fig. 1). Trait research has made a tremendous progress in plant research in the last decade, measuring anything from leaf nutrient contents to drought resistance (e.g. Cornelissen et al. 2003). Thus, trait databases have been compiled on the global level TRY (Kattge et al. 2011) but also with a focus on Germany and certain traits, such as Biolflor (Klotz et al. 2002, Kühn et al. 2004), Leda (Kleyer et al. 2008), Biopop (Poschlod et al. 2003). In contrast, animal trait data has been concentrated mainly on body sizes, but more and more includes also traits related to the trophic position of animals. Environmental data can comprise everything, from climate to soil data as well as data on disturbance regime, temporal stability or landscape structure. Typical LQ questions are how frequent are certain trait values in certain communities, such as what is the proportion of C4 species in the local flora or what is the body size

distribution of animals in different lakes. Typical LR questions comprise species distribution models of which environmental factors limit the range of a particular species as well as temporal changes in species abundances (e.g. Jandt et al. 2011).

However, the aim of understanding general patterns across different floras and faunas requires the elimination of taxon-specific impacts on the analyses. This is the case in RLQ question, where traits are linked to the environment by multiplication of the R, L and Q matrix (e.g. Kröber et al. 2012). RLQ links have also become central to evolutionary research, allowing to ask questions on whether closely related species are more ecologically similar than expected by chance (Mouquet et al. 2012). Combining co-occurrence data with supplementary data also offers the opportunity to assess the relative role of exogenous factors that act on a community compared to interspecific interactions (Kraft et al. 2008).

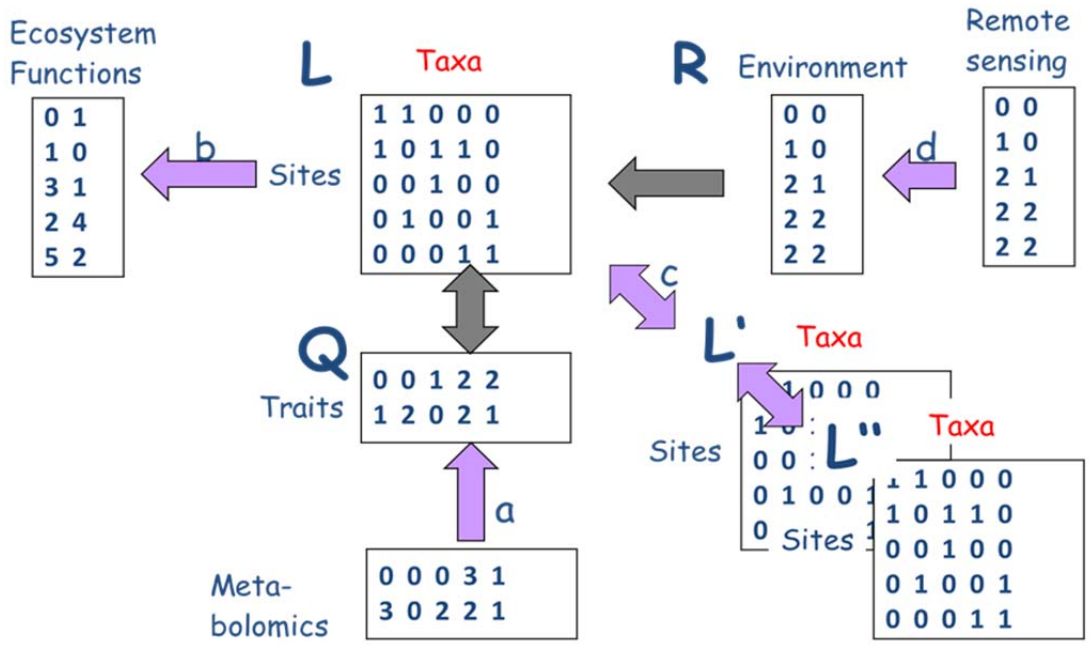

Figure 1: The central role of taxonomic information in biodiversity research. Taxa form the basic unit in occurrence observations, as shown in the central site x taxa L matrix. Taxa are also described by traits (Q matrix) while sites are described by environmental data (R matrix). These three matrices are combined in RLQ questions (gray arrows). Recent developments show further links (purple arrows) to a) metabolomics of taxa, b) ecosystem functioning of sites, c) and remote sensing information on the environment. In addition, d) in multidiversity studies different site x taxa matrices are compared across different trophic levels (denoted as L', L'').

It is clear that linking information from different sources requires a common taxon definition, a requirement that is painfully experienced by everybody who has tried to collate trait information from different sources.

More recently, BD has gone far beyond typical RLQ questions. With the advent of laboratory equipment, such as High-Performance Liquid Chromatography (HPLC), in combination with Diode Array Detectors (DAD), Evaporative Light Scattering Detectors (ELSD), Mass Spectrometry (MS), and Nuclear Magnetic Resonance (NMR), metabolite profiles of species can be obtained across a wide range of taxa and environmental conditions. Similarly, gene expression studies and RNA-Seq libraries using

next generation sequencing (NGS) have resulted in tremendous amounts of data on metabolomics. The challenge is now to aggregate this information on taxon level, to allow for inter-taxa comparisons (arrow "a" in Fig. 1). Again, this can only be done with clear taxon definitions. In contrast to classical traits, metabolomics have been often investigated at below-species levels, i.e. comparing different genotypes, ecotypes or provenances of a single species. This has resulted in the additional challenge of constructing below-species taxonomies (ref.).

Another recent focus in BD research is on the consequences of biodiversity for ecosystem functioning (arrow "b" in Fig. 1). Positive effects of biodiversity on the functioning of ecosystems have been observed in numerous experiments manipulating species diversity of target ecosystems (Loreau *et al.* 2002; Worm *et al.* 2006; Duffy 2009). Such Biodiversity–Ecosystem Functioning (BEF) experiments started with microcosms in the laboratory and with small plots of grassland ecosystems, but today also comprise even forests (e.g. Bruelheide et al. 2013). In principle, BEF questions are currently asked at all taxonomic levels, also on the level of genetic diversity as lower genetic diversity might lead to decreased ecosystem performance, similar as decreased species diversity does (Henery 2011). This is essentially true for all aspects of BD research. Using different genotypes and provenances in BD research is fully equivalent to using different species (as e.g. in invasive species research).

Then, it has become clear that the BD of different trophic levels might be interrelated. This has been shown in BEF experiments where the BD of one trophic level (e.g. plant diversity as in the Jena grassland experiment) has been manipulated with strong consequences for the BD at higher trophic levels (Scherber et al. 2010). Such observations have underpinned the importance of species interactions across trophic levels, and probably will instigate a completely new generation of BEF-experiments. In Fig. 1 this line of research comprises all approaches of community comparisons (arrow "c" in Fig. 1), also those that rely on comparative approaches of community composition across different trophic levels, such as between plant and microbial communities (e.g. Wu et al. 2012).

Finally, our knowledge on the environmental conditions of sites where taxa occur is growing exponentially, mainly through new remote sensing platforms (arrow "d" in Fig. 1). Information from different sensors is becoming increasingly available, as more and more systems are linked, e.g. by the Global Earth Observation System of Systems (GEOSS 2012). Combining passive sensors (delivering typical vegetation characteristics, such as Normalized Difference Vegetation Index, NDVI) and active sensors, such as Light Detection And Ranging (LIDAR) allows for simultaneous measurements of vegetation structure and metabolic processes. This type of information is only linked to the sites, where the taxa occur, not to the taxa themselves, thus is not directly dependent on taxonomical issues. Nevertheless, environmental data require similar considerations as taxon observations. Actually, the increase in ecological information is paralleled in the increase on sequence data in molecular phylogenies. At all levels, data standards have to be developed, and taxon standards is only one single but central aspect of this development. However, as the diversity of data types is steadily increasing also in the field of taxonomy, data standardization and naming conventions are no longer sufficient. Instead, ancillary metadata as used in the Ecological Metadata Language (EML) format have become essential to BD research (Hernández-Ernst et al. 2008).

What are the taxonomic challenges in your research area?

In order to be able to link information in BD research as described in Fig. 1, standard primary data structures, metadata standards, exchange formats, data registration and archiving practices have to be established. Such links can only be established if also the different taxonomic concepts and ecological data concepts are synchronized. In taxonomy, this problem has since long been recognized and resulted in taxon concepts (name-reference couplets), allowing for semantic mediation of taxa, as one name can apply to multiple taxa and a single taxon can have multiple names (Berendsohn 1995). When different projects are included in a database and existing databases are collated, different 'taxonomic concepts' (sensu Geoffroy & Berendsohn 2003) have to be merged. To handle such different concepts, a comprehensive information model for the recording of taxonomic data has been devised for the Global Plant Checklist database project of the International Organisation of Plant Information (IOPI) (Berendsohn 1997). However, the largest challenge is not only the initial establishment of these links but to keep the links updated with time. It is not only taxonomic standards for organisms that vary with time, place, and investigator but also the standards for traits, ecosystem functioning and environmental variables. In this case, the integrity of established relationships as those shown in Fig. 1 will become compromised.

My central proposition here is to base all types of BD information on the approach of concept synonymy. Thus, the taxon concept approach might be enlarged to all other types of BD data. This ranges from the simple problem of joining different cover-abundance scales on which vegetation has been recorded (e.g. Knollová et al. 2005) to different methods of how soil pH values can be measured, from habitat classification approaches (CORINE Habitats Classification, CORINE Landcover, EUNIS 2012) or vegetation types (Rennwald 2000). The development of such concept synonymy data models in other parts of BD research is currently out of reach.

What needs do you see in the next future for your research area?

First of all, taxonomic research has to be made aware of all the different fields of research that are currently linked to species taxonomy. Then, the phenotypic description of taxa lags far behind the describtion of OTUs based on DNA sequences (see contribution of Susanne Renner in this volume). The sequencing of organisms, either of target gene regions or even of the whole genome, is no longer a limiting factor. What is lacking now is to provide phenotypic descriptions such as typical processes mediated by the organisms investigated as well as clear ecological descriptions to which environments the observations apply. Finally, tools have to be developed to keep these databases updated when concepts of taxonomic or other entities change in this database or another. No software is presently existing to integrate different concepts at the various levels of information from plant taxonomy to all other data types of taxon-related information.

Meeting this challenge needs active research. Active research also needs a more pro-active role of taxonomy than presently taken. It is time that taxomists discover that their knowledge how to handle concepts is also urgently needed in many other fields that make use of taxonomic concepts. From my point of view, many other fields in BD science would strongly benefit from lessons taught by taxonomists. Thus, training taxonomists in other fields of BD science and involving taxonomists in ecosystems and ecoinformatic research should be encouraged.

Why taxonomy is relevant to society

Human society needs ecosystems to provide multiple services effectively, especially as human use of natural resources is increasing dramatically, due to population increase and higher per capita consumption rates (Kareiva et al. 2007). The dramatic human-induced loss of BD will compromise the services society receives from nature (Pereira et al. 2010). As taxonomy is an inherent basis in virtually all fields of BD research, taxonomy is also relevant to the services provided by BD. Another important relevance to society is that taxonomy provides names, also common names, to common people. Today's average citizen has a very restricted knowledge on plant and animal species, confined to a few dozen species in total. As BD conservation can only be effective if supported by the majority of the human population, expanding the knowledge on species in schools and universities will be a crucial point in safeguarding BD in the long run.

References

Berendsohn, W.G. (1995): The concept of "potential taxa" in databases. Taxon 44: 207-212.

Berendsohn, W.G. (1997): A taxonomic information model for botanical databases: the IOPI Model. Taxon 46: 283-309.

Bruelheide, H., Nadrowski, K., Assmann, T., Bauhus, J., Both, S., Buscot, F., Chen, X.-Y., Ding, B., Durka, W., Erfmeier, A., Gutknecht, J.L.M., Guo, D.L., Guo, L.-D., Härdtle, W., He, J.-S., Klein, A.-M., Kühn, P., Liang, P., Liu, X., Michalski, S., Niklaus, P.A., Pei, K.Q., Scherer-Lorenzen, M., Scholten, T., Schuldt, A., Seidler, G., Trogisch, S. von Oheimb, G., Welk, E., Wirth, C., Wubet, T., Yang, X.F., Yu, M.J., Zhang, S.R., Zhou, H.Z., Fischer, M., Ma, K.P., Schmid, B. (2013): Designing forest biodiversity experiments: general considerations illustrated by a new large experiment in subtropical China. Methods in Ecology and Evolution 5: 74-89.

CDH (2012) Chorology Database Halle. Available at: http://www2.biologie.uni-halle.de/bot/ag_chorologie/choro/. Last accessed November 2012.

Cornelissen, H., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., et al. (2003) A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Australian Journal of Botany 51: 335–380.

CVH (2012) Chinese Virtual Herbarium. Available at: http://www.cvh.org.cn. Last accessed November 2012.

Duffy, J.E. (2009) Why biodiversity is important to the functioning of real-world ecosystems. Frontiers in Ecology and the Environment, 7, 437-444.

EUNIS (2012) European Environment Agency. Available at http://eunis.eea.europa.eu/. Last accessed November 2012.

Fischer, M., Bossdorf, O., Gockel, S. et al. (2010) Implementing large-scale and long-term functional biodiversity research: the Biodiversity Exploratories. Basic and Applied Ecology 11: 473–485.

Geoffroy, M. & Berendsohn, W.G. (2003): The concept problem in taxonomy: importance, components, approaches. Schriftenreihe für Vegetationskunde 39: 5–14.

GEOSS (2012) Global Earth Observation System of Systems. Available at http://www.earthobservations.org. Last accessed November 2012.

GVRD (2012) German Vegetation Reference Database. Available at: http://www2.biologie.uni-halle.de/bot/vegetation_db/. Last accessed November 2012.

GBIF (2012) Global Biodiversity Information Facility. Available at http://data.gbif.org. Last accessed November 2012.

Henery, M. (2011) The constraints of selecting for insect resistance in plantation trees. Agricultural and Forest Entomology, 13, 111-120.

Hernández-Ernst, V., Poigné, A., Voss, A., Voss, H., Berendsohn, W., Giddy, J., Gebhardt, M., Hardisty, A., Schentz, H. & Magagna, B. (2008) Data & Modelling Tool Structures. Status Report on Infrastructures for Biodiversity Research. Fraunhofer IAIS, Cardiff University.

iDiv (2012) German Centre for Integrative Biodiversity Reserach. Mission. http://www.idiv-biodiversity.de/. Last accessed November 2012.

Jandt, U., von Wehrden, H. & Bruelheide, H. (2011): Exploring large vegetation databases to detect temporal trends in species occurrences. Journal of Vegetation Science 22: 957-972.

Kattge J, Díaz S, Lavorel S, Prentice IC et al. (2011) TRY – a global database of plant traits. Global Change Biology 17:2905–2935.

Kareiva P, Watts S, McDonald R, Boucher T. (2007) Domesticated nature: shaping landscapes and ecosystems for human welfare. Science 316: 1866–1869.

Kleyer, M., Bekker, R.M., Knevel, I.C., Bakker, J.P., Thompson, K., Sonnenschein, M., Poschlod, P., van Groenendael, J.M., Klimes, L., Klimesová, J., Klotz, S., Rusch, G.M., Hermy, M., Adriaens, D., Boedeltje, G., Bossuyt, B., Dannemann, A., Endels, P., Götzenberger, L., Hodgson, J.G., Jackel, A.-K., Kühn, I., Kunzmann, D., Ozinga, W.A., Römermann, C., Stadler, M., Schlegelmilch, J., Steendam, H.J., Tackenberg, O., Wilmann, B., Cornelissen, J.H.C., Eriksson, O., Garnier, E. & Peco, B. (2008): The LEDA Traitbase: a database of life-history traits of the Northwest European flora. Journal of Ecology 96: 1266–1274.

Klotz, S., Kühn, I. & Durka, W. (ed.) (2002): BIOLFLOR–Eine Datenbank zu biologisch-ökologischen Merkmalen der Gefäßpflanzen in Deutschland. Schriftenr. Vegetationskd. 38.

Knollová , I., Chytrý , M., Tichý , L. & Hajek, O. 2005. Stratified resampling of phytosociological databases: some strategies for obtaining more representative data sets for classification studies. Journal of Vegetation Science 16: 479–486.

Kraft, N.J.B., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2008): Trait evolution, community assembly, and the phylogenetic structure of ecological communities. Am. Nat. 170: 271-283.

Kröber, W., Böhnke, M., Welk, E., Wirth, C. & Bruelheide, H. (2012): Leaf trait-environment relationships in a subtropical broadleaved forest in South-East China. PLoS ONE 7: e35742.

Kühn, I., Durka, W. & Klotz, S. (2004): BiolFlor - a new plant-trait database as a tool for plant invasion ecology. Diversity and Distributions 10:363-365.

Loreau, M., Naeem, S. & Inchausti, P. (2002) *Biodiversity and ecosystem functioning.* Oxford University Press, New York.

Michener, W.K. & Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. Trends in Ecology & Evolution 27: 85-93.

Mouquet N., Devictor V., Meynard C. N., Munoz F., Bersier L.-F., Chave J., Couteron P., Dalecky A., Fontaine C., Gravel D., Hardy O. J., Jabot F., Lavergne S., Leibold M., Mouillot D., Munkemuller T., Pavoine S., Prinzing A., Rodrigues A. S., Rohr R. P., Thebault E. & Thuiller W. (2012) Ecophylogenetics: advances and perspectives. Biological Reviews. doi: 10.1111/j.1469-185X.2012.00224.x.

Öpik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davison, J., Kalwij, J.M., Reier, Ü. & Zobel, M. (2010) The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). New Phytologist, 188, 223–241.

NCBI (2012) Nucleotide Database of the National Center for Biotechnology Information. Available at: http://www.ncbi.nlm.nih.gov/nuccore. Last accessed November 2012.

Pereira HM, Leadley PW, Proenca V, et al. (2010) Scenarios for global biodiversity in the 21st century. Science 330: 1496–1501.

Poschlod, P., Kleyer, M., Jackel, A.-K., Dannemann, A. & Tackenberg, O. (2003): BIOPOP – a database of plant traits and internet application for nature conservation. Folia Geobotanica 38: 263–271.

Rennwald, E. (Bearb.) (2000): Verzeichnis und Rote Liste der Pflanzengesellschaften Deutschlands. Schriftenr. Vegetationsk. 35: 1-800.

Schaminée, J.H.J., Hennekens, S.M., Chytry, M. & Rodwell, J.S. (2009): Vegetation-plot data and databases in Europe: an overview. Preslia 81: 173-185.

Scherber C, Eisenhauer N, Weisser WW, Schmid B, Voigt W, Fischer M, Schulze ED, Roscher C, Weigelt A, Allan E, Bessler H, Bonkowski M, Buchmann N, Buscot F, Clement LW, Ebeling A, Engels C, Halle S, Kertscher I, Klein AM, Koller R, Konig S, Kowalski E, Kummer V, Kuu A, Lange M, Lauterbach D, Middelhoff C, Migunova VD, Milcu A, Muller R, Partsch S, Petermann JS, Renker C, Rottstock T, Sabais A, Scheu S, Schumacher J, Temperton VM, Tscharntke T. (2010) Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. Nature 468:553–556.

Senckenberg (2012) Zoological collection. Available at http://sesam.senckenberg.de. Last accessed November 2012.

Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B.C., Lotze, H.K., Micheli, F., Palumbi, S.R., Sala, E., Selkoe, K.A., Stachowicz, J.J. & Watson, R. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science*, **314**, 787-790.

Wu, Y.T., Gutknecht, J., Nadrowski, K., Geißler, C., Kühn, P., Scholten, T., Both, S., Erfmeier, A., Böhnke, M., Bruelheide, H., Wubet, T. & Buscot, F. (2012): Relationships between soil microorganisms, plant communities and soil characteristics in Chinese subtropical forests. Ecosystems 15: 624-636.

Taxonomy and –omics data: two sides of a coin or interdisciplinary linkage via databases, networks and technical solutions?

– Birgit Gemeinholzer

Justus-Liebig-University Giessen, Institute of Botany, Systematic Botany Group, Giessen, Germany

Introduction

Taxonomists are arguably the most active annotators of the natural world, collecting, comparing, evaluating, and publishing phenotypic data to describe new species (Deans, Yoder & Balhoff 2012). Hereby, they thoroughly investigate closely related groups of taxa by comparison of a vast variety of characters, which is a tedious but important work to allow for estimates of biodiversity richness and to explore the origin, organization and sustainability of biodiversity (Wheeler *et al.* 2012). Even during the last centuries the focus of taxonomy did not change and is still aiming at identifying taxa. Taxonomy in the modern era changed from the original morphological species description to incorporation of multiple evidences gained by a variety of newly emerging techniques, providing e.g. chemical, molecular, microscopic data. This changed the way taxonomists work, as now taxonomists must evaluate the explanatory power of characters by analyzing and interpreting large amounts of data.

Data exchange in the taxonomic community

Data exchange in the taxonomic community has long been common practice, e.g. concerning herbarium loans. Since 2001, data exchange if further technically supported by the GBIF (Global Biodiversity Information Facility) Data Portal which is linking worldwide distributed biodiversity data providers in a network-like structure to make taxonomic data searchable from a single point of access. Hereby, GBIF's mission is to make the world's biodiversity data freely and universally available via the Internet (www.gbif.org). The GBIF Data Portal allows for complex searches on any taxon, country or dataset or on a combination of these parameters and is extending its services and components continuously. GBIF's informatics infrastructure builds on existing and emerging international data standards and tools and takes an active part in their development, in close collaboration with Biodiversity Information Standards (Taxonomic Database Working Group TDWG). Several additional initiatives compiling and developing tools and services to enhance the taxonomic workflow and data access and availability are GBIF compliant (e.g. The EDIT Platform for Cybertaxonomy, morphDbase, DNA Bank Network, and others).

Data exchange in the –omics community

Omics-data in general is stored in the databases of the INSDCs (The International Nucleotide Sequence Database Collaboration, EMBL-EBI, GenBank and DDBJ) and/or in specialized community databases (e.g. The Arabidopsis Information Resource (TAIR), the Saccharomyces Genome Database (SGD), the Plant Genome Research database (PGRdb), etc.). Hereby, the Genomic Standards Consortium (GSC) has been formed in 2005 to coordinate the standardization of contextual (meta)data from the point of data acquisition to publication and data submission to the INSDC as prerequisite for any

kind of small- to large-scale data comparison and integration. Also in the omics-community, tools and services are continuously being developed and ameliorated to enhance and facilitate the scientists' workflow, data access and availability - especially for the fast emerging large datasets of the Next Generation Sequencing era. Members of the GSC formed a special Genomic Biodiversity Working Group which is allied with the TDWG group to explore the intersections between research in molecular, systematics/taxonomy, ecology and diversity informatics.

Taxonomy and –omics data exchange:

For their daily routine, currently taxonomists rarely do incorporate –omics-information into their proof of evidence and it is very unlikely that complex omics-data integration into taxonomic investigation will become general routine in the future. The amount of information generated by NGS technologies normally by far exceeds the information needed to group and organize organisms in an evolutionary context. Furthermore, the costs for conducting NGS experiments are too high to be universal applicability, the derived data is of limited explanatory power in a taxonomic context and the procedure for data acquisition is too time-consuming and complex to be conducted by a taxonomist.

On the other hand, -scientists working in the field of –omics research also mostly ignore taxonomic information, e.g. correct organism identification is often neglected and it has not become general routine to fulfill minimum standards for taxonomic documentation to allow for re-identification. Without proper documentation of the source organism, voucher deposition in natural history collection, and/or photo documentation of the specimen under investigation of which –omics data has been generated, the potential community overarching surplus is null.

This has been realized by the GSC community and efforts are being made to overcome this problem, e.g. in form of defining MIMARKS (Minimum information about a marker gene sequence) or MIxS (minimum information about any (x) sequence) (Yilmaz *et al.* 2011).The mutual benefit of the –omics researchers to conduct correct voucher documentation and sharing could be the much longer use and citeability of taxonomic publications, while the taxonomists benefit from much higher impact factors of –omics publications. However, (Arzberger *et al.* 2004) as well as (Enke *et al.* 2012) already stated that the development of appropriate reward structures are necessary component for promoting correct data access and sharing practices. Possibly a reward system between the communities has to be developed that provides incentives for data publication and sharing, either with support of the funding agencies or the publishers.

What are the challenges in your research area?

Technically, data interchangeability and interoperability is not the problem as existing databases storing information for one or the other community already have standards implemented which are interoperable. However, currently data exchange between both communities is not common practice, yet. This seems to be an educational problem, as the surplus of interoperability has not been realized by the different communities, yet.

What needs do you see in the next future for your research area?

- Streamlining procedures to establish confidently applied, uniform taxonomic concepts for the classification of organisms by use of –omics data (Bachmann 1998; Will & Rubinoff 2004).

- Providing techniques which allow for ease of species description and identification (e.g. (Hebert *et al.* 2003)

Why taxonomy is relevant to society

Taxonomy allows for estimates of biodiversity richness. Furthermore taxonomists explore the origin, organization and sustainability of biodiversity (Wheeler *et al.* 2012). Hereby, taxonomists define, describe and recognize species as entities in an evolutionary context. In an era of biodiversity changes we do need to monitor changes as the impacts of biodiversity changes upon our society are still unpredictable. For this we do need data which can be compared and re-evaluated in different contextual dimensions as we are uncertain which questions must be answered for the well-being of our society in the future.

References

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P. & Wouters, P. (2004) Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal,* **3,** 135-152.

Bachmann, K. (1998) Species as units of diversity: an outdated concept. *Theory in Biosciences,* **117,** 213-230.

Deans, A.R., Yoder, M.J. & Balhoff, J.P. (2012) Time to change how we describe biodiversity. *Trends in ecology & evolution (Personal edition),* **27,** 78-84.

Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B. & Gemeinholzer, B. (2012) The user's view on biodiversity data sharing — Investigating facts of acceptance and requirements to realize a sustainable use of research data —. *Ecological Informatics,* **11,** 25-33.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* **270,** 313-321.

Wheeler, Q.D., Knapp, S., Stevenson, D.W., Stevenson, J., Blum, S.D., Boom, B.M., Borisy, G.G., Buizer, J.L., De Carvalho, M.R., Cibrian, A., Donoghue, M.J., Doyle, V., Gerson, E.M., Graham, C.H., Graves, P., Graves, S.J., Guralnick, R.P., Hamilton, A.L., Hanken, J., Law, W., Lipscomb, D.L., Lovejoy, T.E., Miller, H., Miller, J.S., Naeem, S., Novacek, M.J., Page, L.M., Platnick, N.I., Porter-Morgan, H., Raven, P.H., Solis, M.A., Valdecasas, A.G., Van Der Leeuw, S., Vasco, A., Vermeulen, N., Vogel, J., Walls, R.L., Wilson, E.O. & Woolley, J.B. (2012) Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity. *Systematics and Biodiversity,* **10,** 1-20.

Will, K.W. & Rubinoff, D. (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics,* **20,** 47-55.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., Gilbert, J.A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B.W., Blaser, M.J., Bonazzi, V., Booth, T., Bork, P., Bushman, F.D., Buttigieg, P.L., Chain, P.S.G., Charlson, E., Costello, E.K., Huot-Creasy, H., Dawyndt, P., DeSantis, T., Fierer, N., Fuhrman, J.A., Gallery,

R.E., Gevers, D., Gibbs, R.A., Gil, I.S., Gonzalez, A., Gordon, J.I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A.L., Kelley, S.T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C.L., Legg, T., Ley, R.E., Lozupone, C.A., Ludwig, W., Lyons, D., Maguire, E., Methe, B.A., Meyer, F., Muegge, B., Nakielny, S., Nelson, K.E., Nemergut, D., Neufeld, J.D., Newbold, L.K., Oliver, A.E., Pace, N.R., Palanisamy, G., Peplies, J., Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D.A., Assunta-Sansone, S., Schloss, P.D., Schriml, L., Sinha, R., Smith, M.I., Sodergren, E., Spor, A., Stombaugh, J., Tiedje, J.M., Ward, D.V., Weinstock, G.M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J.R., Yatsunenko, T. & Glockner, F.O. (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat Biotech,* **29,** 415-420.

Are the Nomenclatural Codes Fit for the Future? A Protistologist's View

– Regine Jahn

Free University of Berlin, Botanical Garden and Botanical Museum Berlin-Dahlem, Berlin, Germany

Introduction

A common concern of today is that numerous organisms are going extinct before they have even been described. Two of the reasons are that there are too few taxonomists, and that the procedures involved in formally naming new species are too slow. For some groups of species - if not typical animals or plants - it is often difficult to know which of the current *Codes* to apply, and different authors take divergent views. Molecular biology (leading to the recognition of OTUs [Operational Taxonomic Units]) and online taxonomic information management systems ("cybertaxonomy") offer ways to accelerate the recognition and description of new species, and these technologies need to be incorporated into the *Codes* in order to meet the future needs of the scientific community. In addition, there is a need for harmonization of the *Codes*, and work towards a single *BioCode* that can apply to all kinds of organisms. A *BioCode* would help avoid further fragmentation of the scientific communities working with different organisms using a potentially ever increasing numbers of special *Codes*.

The nomenclatural codes

Six nomenclatural codes govern the naming of organisms (for references see (David *et al.* 2012):

- the Zoological Code **ICZN** (International Code of Zoological Nomenclature, 4th edition, International Commission on Zoological Nomenclature 1999); http://www.nhm.ac.uk/hosted-sites/iczn/code/;
- the Botanical Code **ICN** (*International Code of Nomenclature for algae, fungi, and plants* or *Melbourne Code* (McNeill et al. 2012); available online at http://www.iapt-taxon.org;
- the Cultivated Plant Code **ICNCP** (*International Code of Nomenclature for Cultivated Plants*, 8th edition, Brickell et al. 2009); http://www.actahort.org/chronica/pdf/sh_10.pdf);
- the Prokaryote Code **ICNP** (formerly Bacteriological Code; *International Code of Nomenclature of Prokaryotes*, Lapage et al. 1992, see Labeda 2000); http://www.ncbi.nlm.nih.gov/books/NBK8817/;
- the Virus Code **ICVCN** *(The International Code of Virus Classification and Nomenclature*, in Virus Taxonomy, ed. King et al. 2011);
- the **PhyloCode** *(International Code of Phylogenetic Nomenclature* or, version 4c, Cantino and Queiroz 2010); http://www.ohio.edu/phylocode/.

Representatives of above codes work together in the *International Committee on Bionomenclature* (**ICB**) (www.bionomenclature.net) which has recently published papers on the *Draft BioCode 2011* (Greuter *et al.* 2011) and on the *biological nomenclature terms* (David *et al.* 2012). ICB is currently supported financially by the IUBS Programme *BioCode* (2010-2012 & 2013-2015).

What is new?

In summer 2012, the International Commission of Zoological Nomenclature decided that nomenclatural actions can be published electronically if certain criteria are met such as registration of the publication, an electronic archive to preserve the work and the ISSN or ISBN associated with the work (International Commission on Zoological Nomenclature, 2012).

Since 1 January 2012 the new Botanical Code allows electronic publication if certain criteria are met as well as diagnoses in Latin or English (Knapp, McNeill & Turland 2011). This is speeding up publication as can be seen in publications in Open Access Journals such as Phytokeys.In addition, theMelbourne Code implemented registration of new names for the fungi. Already in the Tokyo Code (1994) registration had been planned and tested; in the St.Louis Code (2000) it was kicked out of the Botanical Code.

Ambiregnal taxa

In Melbourne 2011, the Botanical Code changed its name from *International Code on **Botanical** Nomenclature* to *International Code on Nomenclature for **algae**, **fungi** and **plants***. This was done to accommodate phylogenetically diverse groups within the traditional botanical nomenclature. But is the name change a solution for the algae? The term algae include polyphyletic organism groups which do photosynthesis but are not typical plants (Charophytes, Chlorophytes, Glaucophytes, Rhodophytes, Bacilliariophytes, Phaeophytes). In other groups, if sister species do not have chloroplasts, they are termed protozoa and are governed by the Zoological Code (i.e. Cryptophytes / Cryptomonads, Chrysophytes / Chrysoflagellates, Dinophytes / Dinoflagellates, Euglenophytes / Euglenids). In addition, the bluegreen algae have been recognized as prokaryotes and can therefore be ruled by the Prokaryote Code (Cyanophytes / Cyanobacteria). If one taxon can be governed by two codes, these organisms are named ambiregnal taxa. The algae and protozoa communities are beginning to realize these problems and are moving together in using the term protists (Pawlowski *et al.* 2012). There have been intentions to implement a Protist Code but since the protists have turned out to be polyphyletic, this code would also not reflect current taxonomy.

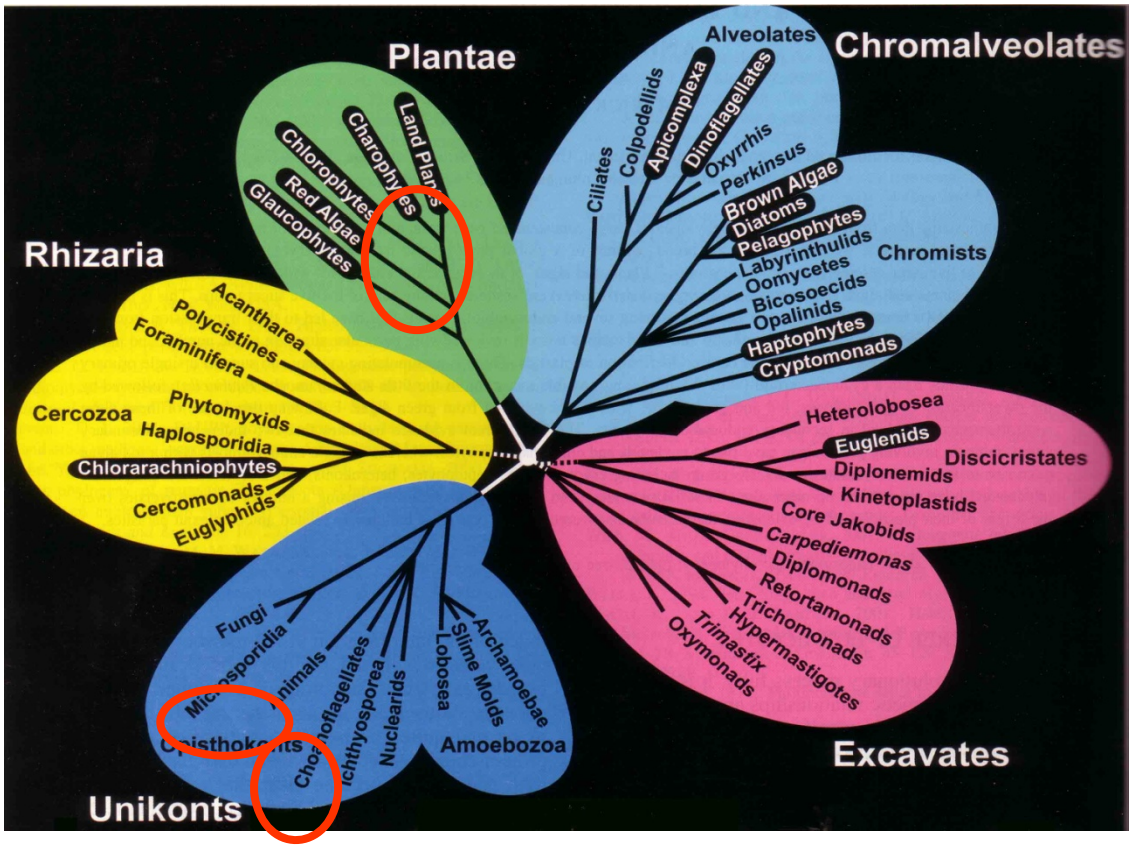

Figure 1: Tree of Eukaryotes and Diversity of Plastid Bearing Eukaryotes (figure taken from Keeling, 2004). Animals, fungi and plants are circled in red; all other groups are considered protists; groups highlighted in black have chloroplasts and are considered algae.

In the end, it should be asked if it is wise to dispose of 200 years of research by dumping old names, types and taxa just for nomenclatural rules. The needs of the users towards nomenclature are stable and unambiguous names i.e. for biomonitoring (EU WFD).

What are the differences?

Under the mandate of the ICB, many terms (such as valid, established, published, legitimate, acceptable) have been harmonized across all codes, facilitating communication in the naming of organisms (see Hawksworth 2010, David *et al.* 2012).

But a number of harmonizations are still needed for ambiregnal taxa. Ambiregnal genera may have different type species under different codes as is the case with *Euglena*'s sister genus *Astasia* which does not have chloroplasts. Besides different names of higher ranks (Euglophyta / Euglenids) and authors, the same genus name has different type species: *Astasia* Dujardin 1841 with its conserved type: *Astasialimpida* Dujardin (ICN) versus *Astasia* Ehrenberg 1830 with its type: *Astasiahaematodes* Ehrenberg (ICZN). Infraspecific ranks (i.e. *Euglenaintermedia* var. *klebsii*), figures as type, epitypification, second step lectotypification are supported by the ICN but not by ICZN. In contrast, homonymy of genus and species is supported by ICZN (*Astasiaastasia*) but not by ICN. Since the bluegreens are now ruled by the Prokaryote Code (Cyanophytes / Cyanobacteria) which requires obligatory registra-

tion and deposition of two living cultures, it is unclear what happens with the names and taxa which had been described before this regulation was implemented.

Homonyms

One of the major clashes between the codes are homonyms. The ICZN in the preamble states: "The objects of the Code are to promote stability and universality in the scientific names of animals and to ensure that the name of each taxon is unique and distinct. All its provisions and recommendations are subservient to those ends and none restricts the freedom of taxonomic thought or actions." But later, it is stated: "Art. 52.7. Homonymy with names of taxa which are not animals. The name of an animal taxon identical with the name of a taxon which has never been treated as animal is not a homonym for the purposes of zoological nomenclature." This situation is similar in the Botanical Code; only the Prokaryote Code is aware of Cross Code Homonyms.

Plantae - Bacillariophyta - Bacillariophyceae - Naviculales - Naviculaceae – ***Navicula***
Animalia - Mollusca - Gastropoda – ***Navicula***

Plantae - Chlorophyta - Chlorophyceae - Chlorococcales - Scenedesmaceae – Didymocystis ***Didymocystis inermis***
Animalia - Platyhelminthes - Trematoda - Azygiida - Didymozoidae – Didymocystis ***Didymocystis inermis inermis***

Plantae - Chlorophyta - Chlorophyceae - Chlorococcales - Scenedesmaceae – ***Lauterborniella***
Animalia - Arthropoda - Insecta - Diptera - Chironomidae – ***Lauterborniella***

Plantae - Bacillariophyta - Bacillariophyceae - Eunotiales - Eunotiaceae – ***Actinella***
Plantae - Magnoliophyta - Magnoliopsida - Asterales - Asteraceae – ***Actinella*** (this name is rejected in the Botanical Code because of intra code homonymy)
Animalia - Mollusca - Gastropoda - Stylommatophora - Helicidae – ***Actinella***

Figure 2: Some examples of homonymy from GBIF (www.gbif.net)

Higher Ranks

The situation with the higher ranks is even more complicated but essential for scientists working with ambiregnal taxa. Within the Botanical Code there are three different endings to be used when classifying fungi, algae or plants. Fig. 3 shows some congruence between botany and zoology (class, suborder and superfamily) but also differences (subclass, family, infrafamily, subtribus). The BioCode would be an instrument to harmonize all organisms into one classification system (see proposal harmonized ranks, Fig.3).

Fig. 3: Comparison of higher ranks between the Botanical and Zoological Code

	Fungi	Algae	Plants	Harmonized ranks	Zoology
Phylum	-mycota	-phyta	-phyta	-biota	-zoa
Subphylum	-mycotina	-phytina	-phytina	-biotina	-zoina
Class	-mycetes	-phyceae	-opsida	-(biot)opsida	-opsida
Subclass	-mycetidae	-phycidae	-idae	-(biot)idae	-zoidae
Order			-ales	-ales	-formes
Suborder			-ineae	-ineae	-ineae
Superfamily			-oidea	-oidea	-oidea
Family			-aceae		-idae
Infrafamily			-idieae		-inae
Subfamily			-oideae		-inae
Tribus			-eae		-ini
Subtribus			-inae		-ina

A plea for the *BioCode*

A unified nomenclatural approach that establishes the BioCode as a forward-looking framework within which nomenclatural problems faced by the previously independent nomenclaturalists can now be discussed in order to harmonize the different special codes, transfer best practices among codes, and to address the problems presented by ambiregnal taxa, was prepared in the *Draft BioCode2011*, (Greuter *et al.* 2011). This *Draft BioCode 2011* received much attention in the nomenclature communities but has now been reduced to the *Framework BioCode,* open for future implementation after harmonization of terms (David *et al.* 2012).

Fit for the Future?

The basis of all major codes is that names need to be attached to types; there can be no conservation without a physical type. This means that molecular data – as long as it is type based – can be incorporated into the diagnosis of a new species. But what about molecular data of taxa which do not have a morphological imprint?

Because of the restrictive rules of its Code in having to deposit cultures for each taxon name, only about 10.000 bacteria species have been described until today. In order to accommodate incomplete descriptions and a taxonomic status for uncultured procaryotic cells which can be recognized by their molecular structures, the prokaryote researchers have invented the term *Putative Taxon* or *Candidatus Status,* which is not a rank but a status which is not formally recognized in the Prokaryote Code (Murray & Stackebrandt 1995). Formal recognition will be done when new observations allow their assignment to a known genus because enough distinguishing characteristics have been recognized. Until today only about 300 *candidatus status* have been assigned to prokaryotes.

The main criticism of this concept is that it is not type based. These questions are raised: Why does molecular data need names? Are we talking about real organisms? Who cares about completing descriptions if naming of OTUs is so easy?

Naming organisms in an age of molecular-based biodiversity discovery

The ICB is currently pursuing the project "Naming organisms in an age of molecular-based biodiversity discovery" within the IUBS programme BioCode (2013-2015). The project will drive forward an international agenda on nomenclatural activities in a world in which discovery of biodiversity is increasingly based on molecular tools. This discovery process is currently not compatible with type-based naming and we are in need of a unified system of naming and registering these new entities and integrating our knowledge of them with that of currently named diversity. This will result in (1) an appendix to the International Committee for Bionomenclature BioCode Framework dealing with policies and guidelines for nomenclature and registration of taxa identified using only molecular identifiers; (2) A new forum for discussion and development of an international agenda on this rapidly developing subject.

References

David, J., Garrity, G., Greuter, W., Hawksworth, D., Jahn, R., Kirk, P., McNeill, J., Michel, E., Knapp, S., Patterson, D., Tindall, B., Todd, J., van Tol, J. & Turland, N. (2012) Biological nomenclature terms for facilitating communication in the naming of organisms. *ZooKeys,* **192,** 67-72.

Greuter, W., Garrity, G., Hawksworth, D.L., Jahn, R., Kirk, P.M., Knapp, S., McNeill, J., Michel, E., Patterson, D.J., Pyle, R. & Tindall, B.J. (2011) Principles and Rules regulating the naming of organisms. *Bionomina,* **3**.

Hawksworth 2010: Terms used in Bionomenclature: the naming of organisms and plant communities. GBIF. http://www.bionomenclature.net/documents/NOMGLOSS_2010_FINAL.pdf.

International Commission on Zoological Nomenclature 2012: Amendment of Articles 8, 9, 10, 21 and 78 of the *International Code of Zoological Nomenclature* to expand and refine methods of publication. ZooKeys **219**: 1–10, doi: 10.3897/zookeys.219.3944.

Knapp, S., McNeill, J. & Turland, N.J. (2011) Changes to publication requirements made at the XVIII International Botanical Congress in Melbourne - what does e-publication mean for you? *PhytoKeys,* **6,** 5-11.

Murray, R.G.E. & Stackebrandt, E. (1995) Taxonomic Note: Implementation of the Provisional Status Candidatus for Incompletely Described Procaryotes. *International Journal of Systematic Bacteriology,* **45,** 186-187.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S.S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A.M., Gile, G.H., Holzmann, M., Jahn, R., Jirku, M., Keeling, P.J., Kostka, M., Kudryavtsev, A., Lara, E., Lukes, J., Mann, D.G., Mitchell, E.A.D., Nitsche, F., Romeralo, M., Saunders, G.W., Simpson, A.G.B., Smirnov, A.V., Spouge, J.L., Stern, R.F., Stoeck, T., Zimmermann, J., Schindel, D. & de Vargas, C. (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *Plos Biology,* **10**.

TAXON-omics – Taxonomie in Deutschland im Zeitalter der „-omics-Forschung“

– Volker Lohrmann

(im Namen der Autoren und Herausgeber der Übersichtsstudie zur taxonomischen Forschung in Deutschland)

Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung an der Humboldt-Universität zu Berlin

Überseemuseum Bremen, Abteilung Entomologie, Bahnhofsplatz 13, 28195 Bremen

Einleitung

Die derzeitige Situation der taxonomischen Forschung in Deutschland wird, wie auch global, durch die Diskrepanz zwischen gestiegenen Anforderungen und Aufgaben sowie den nicht gleichermaßen gestiegenen personellen Mitteln bestimmt (u.a. Lohrmann et al. 2012a, Wheeler et al. 2004). Verstärkt wird diese Situation noch durch teilweise inadäquate Bewertungskriterien der Geldgeber sowie die teils unzureichende Kommunikation des Beitrags taxonomischer Forschung zu gesellschaftlich relevanten Fragestellungen. Dennoch ist kein signifikanter Niedergang der taxonomischen Forschung in Deutschland oder ein Aussterben der Taxonomen anhand von Publikations- oder Artbeschreibungszahlen belegbar (s. Abb. 1 zur Entwicklung der Publikationszahlen taxonomischer Fachartikel; Lohrmann et al. 2012a).

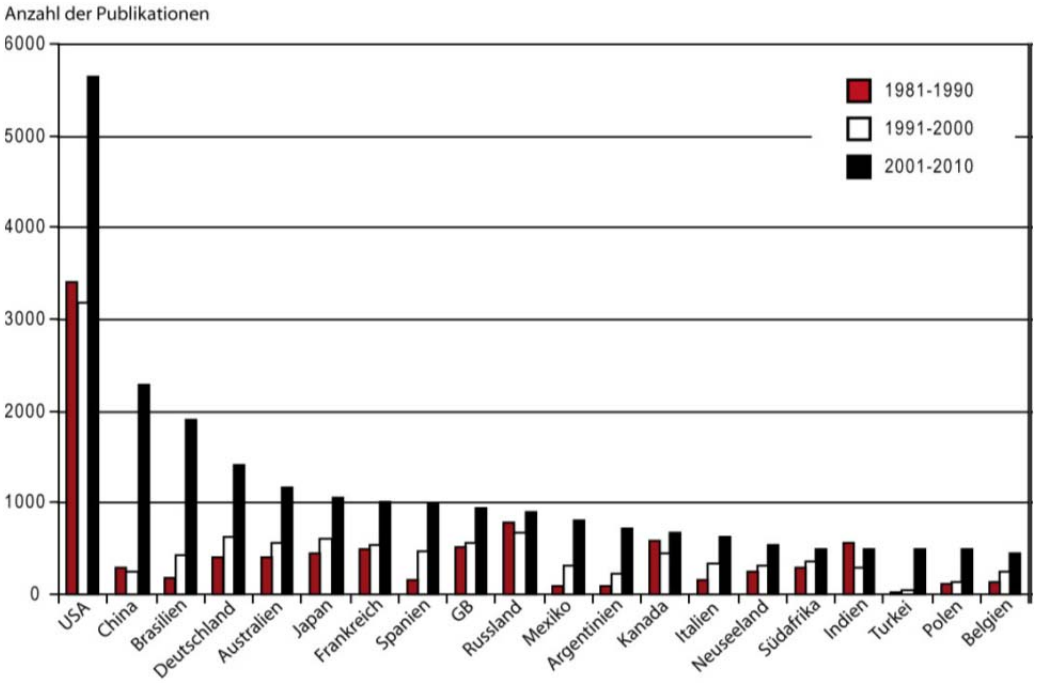

Abbildung 1: Überblick über die Entwicklung der Anzahl an taxonomischen Publikationen, die im Titel dieFormulierung „new species“, „spec. nov.“ oder „n. sp.“ enthalten der „Top 20“-Staaten. Quelle: ISI Web of Science, Thomson Reuters(Abfrage vom 09.06.2011)– Abbildung aus Lohrmann et al. 2012a.

Welches sind die derzeitigen Herausforderungen der Taxonomie?

Die heutige Taxonomie ist genauso wie die allgemeine Wissenschaftslandschaft starken Selektionsdrücken ausgesetzt und damit Teil des gesellschaftlichen Wandels der Wissenschaftsfelder. Dieser Wandel zieht sich mittlerweile wie ein roter Faden durch die taxonomische Forschungslandschaft.

Die heutige taxonomische Forschung ist im Vergleich zu den vorherigen Jahrzehnten integrativer geworden, was sich in einer größeren Vielfalt an Methoden, Theorien und Anwendungsgebieten widerspiegelt (u.a. Dayrat 2005, Lohrmann et al. 2012b). Die mit optischen Hilfsmitteln unterstützten morphologischen Methoden wurden um neue molekulargenetische Methoden ergänzt. Auch die Einbindung taxonomischer Forschung in interdisziplinäre und angewandte Fragestellungen hat ebenso zugenommen wie das erfolgreiche Einwerben größerer Infrastrukturprojekte (s. Abb. 2). Trotzdem mangelt es an personellen Mitteln zur Umsetzung einer modernen taxonomischen Forschung, was signifikant negative Folgen für die Förderung des wissenschaftlichen Nachwuchses sowie für eine angemessene politische und gesellschaftliche Aufmerksamkeit und Förderung hat.

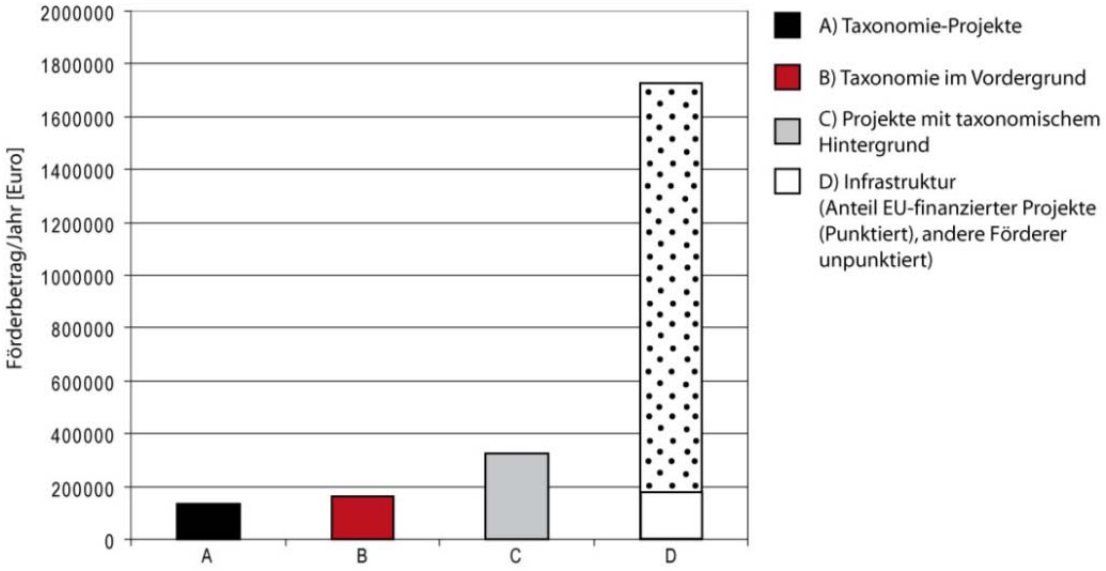

Abbildung 2: Fördersumme taxonomischer Projekte in den entsprechenden Kategorien pro Jahr (N1= 34 Projekte; N2=2.345.345 Euro/Jahr) – Abbildung aus Firtzlaff et al. 2012.

Künftige Anforderungen an die taxonomische Forschung

Auch wenn von politischer Seite dringender Handlungsbedarf besteht, muss sich die Taxonomie ihrerseits der Verantwortung stellen. Es wird dabei künftig weniger darum gehen, die häufig zitierte Kluft zwischen der „antiquierten Taxonomie“ und den „modernen Wissenschaften“ zu überbrücken, als vielmehr darum, die Taxonomie als moderne Wissenschaft zu etablieren – die sie de facto ist. Sieht man von der molekularbiologischen Ausrichtung ab, definiert sich die „-omics-Forschung“ in erster Linie durch ihren ganzheitlichen (holistischen) Ansatz. Dieser ganzheitliche Ansatz kann ebenso auf die moderne taxonomische Forschung übertragen werden (Verwendung aller verfügbaren Merkmalssets zur Artabgrenzung; Molekularbiologie, Morphologie, Ethologie etc.). Ergo: moderne Taxonomie = TAXON-omics.

Einerseits bedeutet dies, dass die existierenden Stärken und der begonnene Wandel (zunehmende Sichtbarkeit, Vernetzung und Teamwork, kaum vergleichbare Methodenvielfalt, große Infrastrukturprojekte und gesellschaftliche Relevanz) der Taxonomie herausgehoben und weiter vorangetrieben werden müssen. Andererseits müssen ebenso die Schwächen (personelle Mittel, Ausbildung, Aufmerksamkeit) erkannt und Maßnahmen zur Verbesserung ergriffen werden. So vermitteln beispielsweise nur noch weniger als 75% der Universitäten, die das Studienfach Biologie anbieten, Arten- bzw. Formenkenntnisse (Chamsai 2012b).

Diesen teils negativen Trends entgegenzuwirken wäre eine wichtige Grundlage, damit die Taxonomie als moderne „-omics“-Forschung wahrgenommen und angemessen gefördert wird.

Durch die Analyse der geographischen Schwerpunkte taxonomischer Forschung in Deutschland hat sich gezeigt, dass diese sich insbesondere dort bündelt, wo die großen naturkundlichen Sammlungen angesiedelt sind (Lohrmann et al. 2012c). Wenn es also darum gehen soll, die taxonomische Kompetenz zu erhalten, wird eine der größten Herausforderungen sein, diese Kompetenzzentren noch stärker und vor allem langfristig für die Ausbildung vom Nachwuchs in die Pflicht zu nehmen. Dies hätte zwar zur Folge, dass eventuell künftig nicht mehr flächendeckend an allen Universitäten gleichermaßen Artenkenntnis gefördert und vermittelt wird (was bereits jetzt der Fall ist), würde aber garantieren, dass sich deutschlandweit einige Leuchttürme im Bezug auf die Vermittlung von taxonomischen Wissen bilden könnten.

Die gesellschaftliche Relevanz der Taxonomie

Die Taxonomie ist eine der Schlüsseldisziplinen der modernen Biodiversitätsforschung und kann neben Grundlagenforschung signifikante Beiträge zu ökonomischen, medizinischen und anderen gesellschaftlich relevanten Problemen leisten (Chamsai 2012a, Smith et al. 2011). Neben der Bereitstellung eines Namens, der Fachübergreifend eine gemeinsame Sprache ermöglicht, ist es die Taxonomie, die durch ihre Expertise entscheidende Beiträge zur Identifikation invasiver Arten, nachhaltigen Bewirtschaftung mariner Ressourcen oder der Entwicklung neuer Medikamente liefert. Zudem ist die Kenntnis der Arten und ihrer Interaktionen nötig, um im politischen Raum die Erreichung – oder auch Nichteinhaltung – von Biodiversitäts- und Nachhaltigkeitszielen zu dokumentieren (Vohland & Lohrmann 2012).

Literatur

Dayrat, B. (2005) Towards integrative taxonomy. – Biological Journal of the Linnean Society 85, 407–415.

Chamsai, L. (2012a) Kapitel 6. Wo wird taxonomische Forschung benötigt? -Anwendungsbereiche. *In*: Lohrmann, V., Vohland, K., Ohl, M. & C. Häuser (Hrsg.) (2012): Taxonomische Forschung in Deutschland - Eine Übersichtsstudie. 122 S. DOI: 10.5075/taxo-studie-01-2012

Chamsai, L. (2012b) Kapitel 8. Ausbildung. *In*: Lohrmann, V., Vohland, K., Ohl, M. & C. Häuser (Hrsg.) (2012): Taxonomische Forschung in Deutschland - Eine Übersichtsstudie. 122 S. DOI: 10.5075/taxo-studie-01-2012

Firtzlaff, V., Lohrmann, V. & K. Vohland (2012) Kapitel 7. Ressourcenallokation. *In*: Lohrmann, V., Vohland, K., Ohl, M. & C. Häuser (Hrsg.) (2012): Taxonomische Forschung in Deutschland - Eine Übersichtsstudie. 122 S. DOI: 10.5075/taxo-studie-01-2012

Lohrmann, V., Chamsai, L., Häuser, C. & K. Vohland (2012a) Kapitel 1. Einleitung in die Studie zur taxonomischen Forschung in Deutschland. *In*: Lohrmann, V., Vohland, K., Ohl, M. & C. Häuser (Hrsg.) (2012): Taxonomische Forschung in Deutschland - Eine Übersichtsstudie. 122 S. DOI: 10.5075/taxo-studie-01-2012

Lohrmann, V., N. Enke & B. Gemeinholzer (2012b) Kapitel 4. Mit welchen Methoden arbeiten Taxonomen? *In*: Lohrmann, V., Vohland, K., Ohl, M. & C. Häuser (Hrsg.) (2012): Taxonomische Forschung in Deutschland - Eine Übersichtsstudie. 122 S. DOI: 10.5075/taxo-studie-01-2012

Lohrmann, V., C. Quaisser, L. Chamsai, C. Häuser & K. Vohland (2012c) Kapitel 3. Sammlungen als Forschungsinfrastruktur. *In*: Lohrmann, V., Vohland, K., Ohl, M. & C. Häuser (Hrsg.) (2012): Taxonomische Forschung in Deutschland - Eine Übersichtsstudie. 122 S. DOI: 10.5075/taxo-studie-01-2012

Smith, R., Rassmann, K., Davies, H. & N. King (eds.), 2011. Why Taxonomy Matters. http://www.bionet-intl.org/why. BioNET-INTERNATIONAL, Egham, UK. (Accessed on 15/12/2012).

Vohland, K. & V. Lohrmann (2012) Kapitel 10. Schlussfolgerungen. *In*: Lohrmann, V., Vohland, K., Ohl, M. & C. Häuser (Hrsg.) (2012): Taxonomische Forschung in Deutschland - Eine Übersichtsstudie. 122 S. DOI: 10.5075/taxo-studie-01-2012

Wheeler Q. D., P. H. Raven and E. O. Wilson (2004) Taxonomy: impediment or expedient? – Science 303, 285.

Main challenges in the taxonomy of land plants (embryophytes)

– Susanne S. Renner
University of Munich (LMU), Botanische Staatssammlung Munich and Botanical Garden Munich, Germany

The utility of electronic images for naming embryophytes, and permanent collections of preserved specimens as a basis for identification via barcodes

Most land plants can be preserved in herbaria, and for most it is possible to make duplicate specimens of the same genetic individual to be kept in multiple herbaria, which is the basis for "spreading" names and species concepts. To identify land plants, one usually compares unnamed specimens to named specimens in public collections. (Very few of the World's countries have good written floras with keys.) Comparing electronic images of unnamed specimens to already named specimens usually suffices to decide whether a name may apply or not, especially since geographic information is usually available for both the names and the unnamed material. Over the past 10 years, this has led to huge imaging projects aiming to place online all type material held in herbaria worldwide (e.g., the Global Plant Initiative: http://gpi.myspecies.info), with higher plants especially suited for imaging since their herbarium specimens are more or less 2-dimensional. Optical character recognition software helps capturing label data from the imaged specimens. The availability of these images has greatly sped up the naming of plant material, especially from tropical countries with high biodiversity and little taxonomic literature. Obviously, not all plants are identifiable via images (think of liverworts, leafy mosses, hornworts, or lichens), but most herbaria now prefer to first offer sending electronic images before sending specimens, especially as regards types.

These digitization efforts have run in parallel to barcoding efforts because **barcoding cannot lead to naming of known or new species unless already named specimens are barcoded**. **The correct naming of herbarium-stored material thus is essential for plant barcoding projects**.

Since efforts to barcode all "life" depend on the maintenance of collections to house the related voucher specimens, frozen tissue, or sometimes also tubes with extracted DNA, the worldwide demand for permanent storage of biological material in biorepositories is increasing, and new such collections are constantly being founded. In mid-2012, this also led to the merger of the online Index Herbariorum (and database of all herbaria and their staff), the Biodiversity Collections Index, and Bioreprositories.org. The merged database now provides information about physical collections plus data held by the Consortium for the Barcode of Life (CBOL) and the National Center for Biotechnology Information (GenBank). Basically, this links voucher specimens in collections to sequence data records in GenBank.

Current flowering of taxonomic research and the persistent challenge from synonymous names

More people are describing more new species than ever before in the history of mankind (Joppa et al. 2011, Costello et al. 2012a, Bacher 2012). There has been a 2.5 fold increase in the proportion of taxonomists in South America and the Asia-Pacific region compared to Europe and North America from the 1980's (Gaston and May 1992) to the 2000's (Costello et al. 2012b), reflecting the increasing number of scientists in developing countries. The number of publications describing new species increased each decade from 1980 to 2010 (Lohrmann et al. 2012). From 2000-2009, one study found that 8,600 people described 30,484 species (Costello et al. 2012a), and another reported that 166,000 species were named (Wheeler and Pennak 2012). **These numbers are evidence that the entry of molecular methods into systematics has not been accompanied by a decline of the rate of description and naming of new species worldwide**. Instead, **molecular data appear to have sped up the recognition of new species, though not necessarily the rate of description and naming of new species.**

Naming new taxa, however, does not solve the huge problem of synonymous names. In land plants, 46% of all land plant names are estimated to be synonyms (The Plant List, http://www.theplantlist.org, but also many previous assessments summarized in Scotland et al. 2003). Clearly, superfluous names lead to wrong inferences and wrong research results.

There is an urgent need for the reduction of synonymous names, which is best accomplished by sequencing and comparing barcode-suitable genome regions from type specimens or specimens collected close to type localities (so-called topo-types).

References

Bacher, S. 2012. Still not enough taxonomists: reply to Joppa *et al.*. *Trends Ecol. Evol.* 27: 65-66.

Costello, M. J., Wilson, S. P., and B. Houlding. 2012a. Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Syst. Biol.* 61: 871-883.

Costello, M. J., Wilson, S. P., and B- Houlding. 2012b. More taxonomists describing significantly less species per unit effort may indicate that most species have been discovered. *Systematic Biology,* in review

Gaston, K. J. and May R. J. 1992. Taxonomy of taxonomists. *Nature* 356: 281-282.

Joppa, L.N., Roberts, D. L., and S. L. Pimm. 2011. The population ecology and social behaviour of taxonomists. *Trends Ecol. Evol.* 26: 551-553.

Lohrmann, V., Vohland, K., Ohl, M., and C. Häuser. 2012. *Taxonomische Forschung in Deutschland – Eine Übersichtsstudie.* Museum für Naturkunde, Berlin.

Scotland, R. W. and A. H. Wortley.2003. How many species of seed plant are there? *Taxon* 52**:** 101-104.

Wheeler, Q. D. and S. Pennak. 2012. *State of Observed Species*. International Institute for Species Exploration, 11pp. Retrieved from http://species.asu.edu/SOS, 20th 182 February 2012.

Phylogenomics versus phylophysiology – static versus dynamic

– Wolfram Weckwerth
University of Vienna, Department of Molecular Systems Biology, Vienna, Austria

Introduction

In the last decades – since the initial publication of the first human and first plant genome in 2000 – bioanalytical platforms have been developed which have revolutionized the molecular description of an organism. For instance, next generation sequencing technology has accelerated the speed of genome sequencing and the characterization of natural genetic variation exponentially. As a consequence, comparative genomics and the analysis of polymorphisms are basic resources for phylogenetic studies nowadays and are complementing classical approaches such as molecular systematics based on single or multiple gene sequences as well as morphological and anatomical studies. Furthermore the active molecular phenotype is measurable at a genome-scale level. Consequently, these molecular studies are nowadays fundamental requirements for the investigation of the phenotypic response of an organism to its specific environment. However, the final linkage between morphological and anatomical data of an organism and these “OMICS”-data is missing.

Taxonomics: A “new” tool in a systems biology framework

Taxonomy provides an in depth classification of organisms based on morphological, anatomical, molecular and other characteristics such as specific environment and habitat descriptions. Although the primary aim of taxonomy is not evolutionary relationships but classification, the separation of phylogenetic characterization of organisms, evolutionary relationships and the classification according to morphological, anatomical and molecular data is almost impossible. Systems biology aims to functionally understand an organism as the result of the integration of many different complex and nonlinear layers of organization at the molecular and morphological level in well-controlled experimental conditions up to natural ecosystems (Weckwerth 2011a). What is nowadays missing in many complex research projects is a thorough morphotypic description of the organism of interest.

Taxonomic research is highly specialized on the morphotypic description to develop unambiguous rules for organism classification. The morphotype, however, even of clones of the same genotype is highly dynamic due to physiological and morphological adaptation strategies resulting in a large range of within-species and ecotypical variation (Hiesey, Clausen & Keck 1942; Laibach 1943; Hoffmann & Sgro 2011). Taxonomic research delivers rules and instruments providing means for this phenotypical or phylophysiological classification. First, taxonomy defines unambiguously distinguished groups of organisms based on very specific characteristics such as flower morphology in angiosperms. Secondly, the range of different morphotypes as a result of environmental adaptation has to be explored. The characterization of these morphotypes is one of the fundamental challenges of taxonomic research and thereby an integral part of organismal and systems biology.

What are the challenges in your research area?

Despite the availability of thousands of genome sequences the prediction of metabolism, physiology or phenotype from the genotype is still impossible (Weckwerth 2011b). The linkage of genotype-data and phenotype-data of an organism in its natural habitat will be the strongest challenge in the next decades of biology to develop rules which allow the design of predictive models based on a genome sequence.

One bottleneck is that phenotype description and species characterization are difficult because of within-species-variation and morphotypic plasticity.

The integration of molecular data such as genomic sequences, polymorphisms and analysis of genome-scale genetic variation by genome-wide association studies (GWAS), transcriptomics, proteomics and metabolomics can address these questions and help to understand the molecular plasticity of an organism as an adaptation process to the environment (Hoffmann & Sgro 2011; Weckwerth 2011a). At the same time this adaptation process shapes the morphology and genotype of the organism thereby leading to different rates of reproduction and processes of natural selection, thus different speciation rates (Weckwerth 2011a). These processes can be investigated for the first time at a very comprehensive genome-wide molecular scale using for instance GWAS and OMICS technologies. To understand, how the genotype shapes the morphotype and environmentally-driven phenotype and vice versa, all the molecular, ecophysiological and phenotypic data need to be integrated.

What needs do you see in the next future for your research area?

One of the strongest needs of the future are databases and data integration. Strong efforts are underway to organize and consolidate knowledge of single model systems such as *Arabidopsis thaliana, Chlamydomonas reinhardtii, Rice, Yeast,* and many more. Each model organism also shows a large diversity of various ecotypes and closely related species, e.g. hundreds of different ecotypes for *Arabidopsis thaliana* allow profound studies of adaptation and speciation processes. All this knowledge is organized in databases and the research community feels responsible to publicly provide these data (Baerenfaller *et al.* 2012). These databases are essential instruments in the future to link taxonomic knowledge, phylogenetic characterization, molecular data and genotype-phenotype-relationship for a better understanding of organismal diversity, adaptation processes of organisms to different environments and the role of morphology, anatomy and natural genetic variation. Databases of functional genome annotation, natural genetic variation and polymorphisms from GWAS, of proteomics data, of metabolomics data for chemosystematics, etc. will play a major role for the interpretation and description of organisms and their biodiversity.

Why taxonomy is relevant to society

Taxonomy provides the backbone for biological research. Without a common classification system we will not be able to exchange and compare worldwide generated scientific data in view of giving functional interpretations of OMICS-data, revealing evolutionary relationships, and understanding natural genetic variation and especially the complex process of speciation. These processes are most relevant for the functioning of ecosystems. If we do not try to understand ecosystem functioning in

more detail and at a global scale we compromise the balance and sustainability of our life on earth (Weckwerth 2011a). Thus, taxonomy and ecology/biodiversity/ecosystem-functioning are intimately linked disciplines which cannot be separated from each other.

References

Baerenfaller, K., Bastow, R., Beynon, J., Brady, S., Brendel, V., Donaldson, S., Dooley, R., Forster, M., Friesner, J., Gifford, D., Grotewold, E., Gutierrez, R., Huala, E., Jaiswal, P., Joshi, H., Kersey, P., Liu, L., Loraine, A., Lyons, E., May, S., Mayer, K., MacLean, D., Meyers, B., Mueller, L., Muller, R., Muller, H.M., Ouellette, F., Pires, J.C., Provart, N., Staiger, D., Stanzione, D., Taylor, J., Taylor, C., Town, C., Toyoda, T., Vaughn, M., Walsh, S., Ware, D., Weckwerth, W. & Informatics, I.A. (2012) Taking the Next Step: Building an Arabidopsis Information Portal. *Plant Cell,* **24,** 2248-2256.

Hiesey, W.M., Clausen, J. & Keck, D.D. (1942) Ecological aspects of evolution - Relations between climate and intraspecific variation in plants. *American Naturalist,* **76,** 5-22.

Hoffmann, A.A. & Sgro, C.M. (2011) Climate change and evolutionary adaptation. *Nature,* **470,** 479-485.

Laibach, F. (1943) Arabidopsis thaliana (L.) Heynh. als Object für genetische und entwicklungsphysiologische Untersuchungen. *Bot. Archiv.,* **44,** 439-455.

Weckwerth, W. (2011a) Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J Proteomics,* **75,** 284-305.

Weckwerth, W. (2011b) Unpredictability of metabolism--the key role of metabolomics science in combination with next-generation genome sequencing. *Anal Bioanal Chem,* **400,** 1967-1978.

Textbeiträge Workshop Mikrobiologie, medizinische Mikrobiologie, Mykologie

Microbiomes and metagenomes in need of professional classification

– Rudolf Amann

Max Planck Institute for Marine Microbiology, Bremen, Germany

Introduction

Although usually invisible to the human eye *Bacteria* and *Archaea* (Woese and Fox 1977) – the two prokaryotic domains of microbial life – account for roughly 50% of the global biomass (Whitman et al. 1998). The physiological diversity of microorganisms is truly amazing, ranging from energy generation by respiration of organic matter with oxygen to oxidation of reduced inorganic compounds like sulphur, ammonia or Fe(II) with sulphate or nitrate, and from anaerobic fermentations to photosynthesis. In particular *Archaea*, but also some *Bacteria* have adapted to life in extreme environments. Well-known examples are hyperthermophilic *Archaea* which grow at temperatures well above 100 ^{o}C and halophilic *Archaea* and *Bacteria* able to dwell in saturated salt brines. From pole to pole *Bacteria* and *Archaea* have colonized literally all environments where they are present in often tremendous numbers. The seemingly transparent water of lakes and oceans contains usually about 1 million *Bacteria* and *Archaea* per millilitre, in surface soils and sediments numbers per gram often exceed 100 million (Whitman et al. 1998).

Considering their sheer abundance of > 10^{30} individuals (Whitman et al. 1998) and the fact that microbial life on planet Earth has continuously evolved for almost 4 billion years into numerous highly diverse niches, the number of validly described species of *Bacteria* and *Archaea* of only ~10.000 (Yarza et al. 2010) can only be the tip of the iceberg. Due to the strict Bacteriological Code requiring the isolation and detailed geno- and phenotypic characterization of pure cultures and their deposition in two culture collections, microbial taxonomists are slow in describing novel species. As a consequence they have in contrast to botanists and zoologists only validly described about 1% of extant bacterial and archaeal species estimated to be in the range of 10^{6} species (Yarza et al. submitted). Due to slow growth, symbiotic relationships with other organisms and highly specialized metabolisms – all three of these problems coincide e.g. in the mixed bacterial-archaeal consortia catalysing the globally relevant process of anaerobic oxidation of methane with sulphate (Boetius et al. 2000) - a major part of the microbial diversity will also in the future not be accessible by traditional cultivation-based approaches (Amann et al. 1995).

These hidden microbial worlds can today be readily accessed by high throughput sequencing of nucleic acids directly extracted from environmental samples. Thereby, “metagenomes” are obtained which are defined as the sum of the genomes of the different organisms present in a habitat. Cultivation-independent, nucleic acid-based studies are also increasingly applied to study the “human microbiomes”, the metagenome of the microorganisms each of us is carrying, e.g., in the gut or on the skin. It has been suggested that these microbiomes, which are modulated by our diets, are hot spots of horizontal gene transfer (Hehemann et al. 2010). Their impacts on human health issues like im-

mune maturation (Chung et al. 2012) and obesity (Turnbaugh et al. 2006) are actively studied. Also in the field of environmental microbiology metagenomics has proven a very powerful tool to assess the composition and functional repertoire of microbial communities. By combining comparative metagenomics with meta-proteomics and the in situ identification and quantification of single cells by fluorescence in situ hybridization (Amann et al. 1995) a substrate-driven succession of discrete bacterioplankton clades could be shown in response to a spring diatom bloom in the German Bight of the North Sea (Teeling et al. 2012), demonstrating biochemical mechanisms in the mineralization of marine polysaccharide that are similar to those in human digestion.

The huge sequence data sets resulting from medical, environmental or biotechnological microbiome and metagenome project are in need of professional classification. The analysis of these giga base pair databases requires the assignment of assembled genome fragments to particular "bins" e.g. by tetranucleotide frequence analysis (Teeling et al. 2004), e.g. together with a simple analysis of the G+C content of genome fragments (Fig. 1). These "bins" reflect the genomes of particular clades of *Bacteria* and *Archaea*.

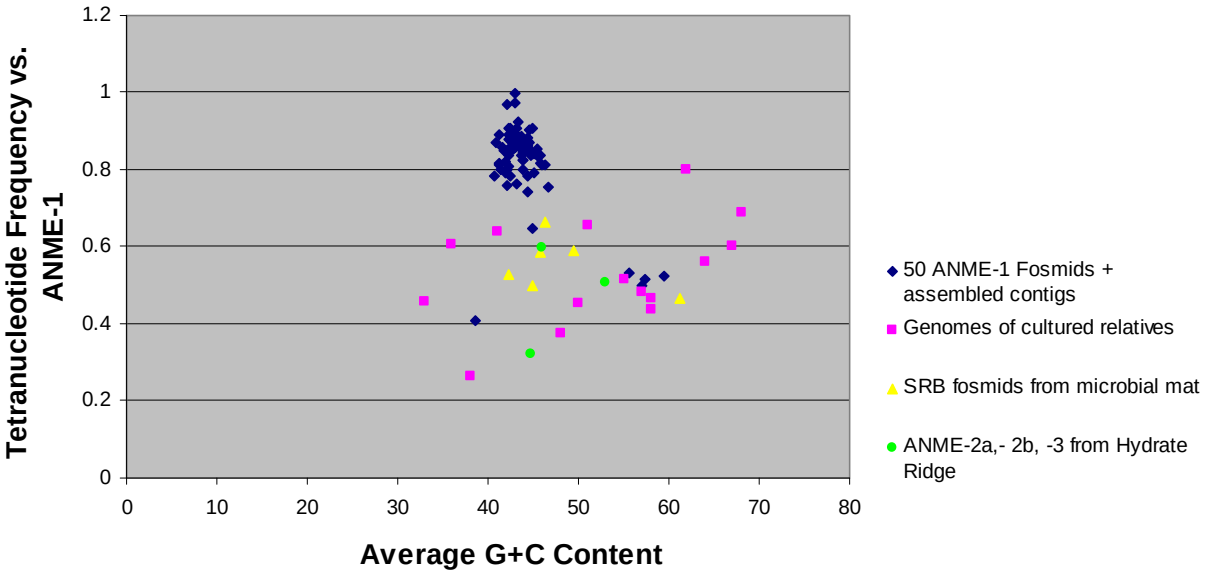

Figure 1: By a combination of average DNA G+C content analysis (x axis) and tetranucleotide frequency analysis (y axis) a tight cluster (= "bin") of 35-40 kbp genome fragments can be separated from a background of unrelated DNA fragments. The blue-labelled fragments in the upper region of the graph likely originate from ANME-1 Archaea involved in the anaerobic oxidation of methane with sulphate (Meyerdierks et al. 2010)

Taxonomists can contribute to the analysis of microbiomes and metagenomes by building and maintaining curated data bases of fully described pure cultures (e.g. Yarza et al. 2010), but also by a preliminary classification of the numerous yet uncultured species that extends beyond sequence information only (Murray and Schleifer, 1994). It would be highly advantageous if this classification in candidate taxonomic units would be done in a standardized way which is convergent with currently practiced boundaries of bacterial and archaeal genera, families, orders or classes (Yarza et al. submit-

ted). A first interpretation of the size of such a metagenomic "bin" is often done by comparative sequence analysis of phylogenetic markers like 16S rRNA genes.

The classification of microbiomes and metagenomes are today highly relevant in disciplines as different as medicine, biotechnology, agriculture and environmental sciences. Each genomic fragment contains functional as well as taxonomic information. It will therefore require the input of trained taxonomist to analyze microbiome and metagenomes and integrate this knowledge into a more holistic picture of biodiversity.

Literature

Amann, R.I., Ludwig, W., and Schleifer, K.-H.:Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. *59*, 143-169 (1995).

Boetius, A., Ravenschlag, K., Schubert, C., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jørgensen, B.B., Witte, U., and Pfannkuche, O.: A marine microbial consortium apparently mediating anaerobic oxidation of methane. Nature *407*, 623-626 (2000).

Chung, H., Pamp, S.J., Hill, J.A., Surana, N.K., Edelman, S.M., Troy, E.B., Reading, N.C., Villablanca, E.J., Wang, S., Mora, J.R., Umesaki, Y., Mathis, D., Benoist, C., Relman, D.A., and Kasper, D.L.: Gut immune maturation depends on colonization with a host-specific microbiota. Cell 149, 1578-1593 (2012).

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G: Transfer of carbohydrate active enzymes from marine bacteria to Japanese gut microbiota. Nature *464,* 908-914 (2010).

Meyerdierks, A., Kube, M., Kostadinov, I., Glöckner, F. O., Reinhardt, R., and Amann, R.: Metagenome and mRNA expression analyses of anaerobic archaea of the ANME-1 group. Environmental Microbiology *12,* 422-439 (2010).

Murray, R.G.E., and Schleifer, K.H.: Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes. Int. J. System. Bacteriol. *44,* 174-176 (1994).

Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., and Glöckner, F.O:. Application of tetranucleotide frequencies for the assignment of genomic fragments. Environ. Microbiol. *6,*938-947 (2004).

Teeling, H, Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., Kassabgy, M., Huang, S., Mann, A.J., Waldmann, J., Weber,M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F.D., Callies, U., Gerdts, G., Wichels, A., Wiltshire, K.H., Glöckner, F.O., Schweder, T., and Amann, R.: Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science 336, 608-611 (2012).

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and gordon, J.I.: An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027-1033 (2006).

Whitman, W. B., Coleman, D. C., and Wiebe, W. J.: Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. USA *95*, 6578-6583 (1998).

Woese, C.R., and Fox, G.E.: Phylogenetic structure of the prokaryotic domain – primary kingdoms. Proc. Natl. Acad. Sci. USA *74*, 5088-5090 (1977).

Yarza, P., Ludwig, W., Euzeby, J., Amann, R., Schleifer, K.-H., Glöckner, F.O., and Rossello-Mora, R.: Update of the all-species Living Tree Project based on 16S and 23S rRNA sequence analyses. System. Appl. Microbiol. 33, 291-299 (2010).

Yarza, P., Yilmaz, P., Prüße, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzeby, J., Amann, R., and Rossello-Mora, R.: Reconciling taxonomic classification of Archaea, Bacteria and their environmental diversity by means of SSU rRNA gene sequences. (submitted to Nat. Microbiol.Rev.).

Fungal genomics and its impact on secondary metabolite discovery and elucidation of pathogenicity

– Axel A. Brakhage

Leibniz-Institute for Natural Product Research and Infection Biology (HKI), Department of Molecular and Applied Microbiology, Jena, Germany

Friedrich-Schiller-University Jena, Chair of Microbiology and Molecular Biology für Mikrobiologie, Jena, Germany

Fungi are more closely related to animals than many other eukaryotic organisms. These two kingdoms diverged from their last common ancestor, which was proposed to be a unicellular organism that lived in the oceans propelled by a flagellum, on the order of a billion years ago. Of the estimated 1.5 million fungal species that exist today, about 130,000 are known and more than 200 species have been associated with humans, either as pathogens or as commensal organisms. Over time, there will be an increase in the number of fungal pathogens of humans due to the aggressiveness, the consequences of modern medical care. Fungi that infect humans have evolved several times independently. Most plant diseases are caused by fungi making these organisms very important in agriculture. In addition, fungi are invaluable resources. Consider the usefulness of the many secondary metabolites fungi produce such as life-saving antibiotics including penicillins and cephalosporin, as well as drugs that facilitate organ transplants (cyclosporine) and that reduce cholesterol (statins), and insecticides. Many fungi have found use in biotechnology.

Theses

- The species concept of fungi is outdated. Prior to the advent of molecular biology, certain fungi were assigned two names: one for the sexual stage (teleomorph) and another for an asexual reproductive stage (anamorph). This convention is an anachronism that dates from a period in microbiology when, in the absence of molecular evidence, fungi were classified solely on the basis of their morphological form. Dual nomenclature confuses fungal systematics. A species designation should accurately convey the essential characteristics that identify an organism as a member of a particular species.

- Members of a fungal species can produce entirely different compounds. Does this define individual lineages within one species? How does chemical diversity of fungi relates to the species concept?

- We do not know, how many fungal clones (genotypes/isolates) are present in an infected patient.

- Comparative genomics reveals insight in virulence traits of fungal pathogens.

Fungal biodiversity with respect to carbon utilisation: Taxonomic relationships versus parallel evolution

– Ronald P. de Vries

CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands

Introduction

Fungi can grow on many different carbon sources, but plant biomass is the main carbon source for a majority of the fungi. Three main lifestyles are found in relation to this substrate: saprobe, pathogen and symbiont. These different lifestyles require different approaches in how to interact with their substrate, since symbionts depend on the plant to be healthy, while pathogens often kill their host. Due to this and the variations in plant biomass in different biotopes, the approach of fungi to degrade or modify plant biomass differs strongly. Adaptation to their biotope will therefore push for strong evolution of the part of the genome that is related to carbon source utilization, which may lead to differences that are much larger than the general taxonomic differences between species.

Degradation of plant polysaccharides by fungi

Polysaccharides are the main component of plant biomass and also the main substrate for most fungi that consume plant biomass (de Vries & Visser 2001). However, fungi cannot take up these polymeric compounds, but need to first degrade them to their monomeric and small oligomeric components using large sets of extracellular enzymes. The production of these enzymes needs to be tightly controlled to be a good match for the available substrate, so that no energy is wasted on the production of non-relevant enzymes (de Vries 2003). The current model on how fungi do this is depicted in Fig. 1. The presence of a monomeric component of a polysaccharide is sensed by the fungus, either by cell-wall bound sensors or by aspecific import of the compound followed by intracellular sensing. This then starts a signal-transduction pathway, which ends with the activation of a transcription regulator. This regulator gets transported into the nucleus and binds to the promoter of its target genes. In most cases, these target genes include both genes to metabolize the compounds as well as genes that encode extracellular enzymes that can degrade the polysaccharide the compound most likely originated from. This will result in the release of more of that compound, which keeps the circle going. Several regulators involved in this process have been identified in a small number of fungi, especially in *Aspergillus* (de Vries 2003; Battaglia *et al.* 2011; Gruben 2012; Vankuyk *et al.* 2012), while the genes encoding metabolic and extracellular enzymes are more commonly known for fungi (Battaglia *et al.* 2010; van den Brink & de Vries 2011). The way in which the fungus senses the compound as well as the signal transduction pathways involved are still largely a black box.

Taxonomy versus parallel evolution

While it is often believed that performing taxonomy/phylogeny on whole genomes will improve fungal systematics, the results obtained in genes related to carbon utilization indicate that this may not be the case. As mentioned above, the available carbon sources in a biotope create a very string evolutionary selection on the genes related to converting them and genome analysis of several fungi has

demonstrated a good match between genome content and biotope with respect to carbon utilization. One of the clearest examples of this is the ascomycete *Podospora anserine*, which only lived as a late colonizer in herbivore dung. In this biotope the main plant-based substrate is lignocellulose, a mixture of cellulose, xylan and lignin. In contrast, species from the genus *Aspergillus* are commonly found in nearly every biotope and are therefore exposed to a much wider variety of substrates. Comparing the genomes of these fungi, demonstrated that the genome of *P. anserine* is strongly enriched in genes encoding cellulolytic and xylanolytic enzymes but poor in genes related to the degradation of other polysaccharides, such as pectin, mannan and inulin (Espagne *et al.* 2008; Coutinho *et al.* 2009). The genomes of the Aspergilli have genes encoding a much broader set of plant polysaccharide degrading enzymes, but lower numbers of cellulolytic and xylanolytic genes. Interestingly, the potential for plant biomass degradation of *P. anserine* is more similar to that of basidiomycete fungi than to most other ascomycete fungi (Espagne *et al.* 2008). This implies that selection pressure based on biotope has a stronger effect on the evolution of the part of the genome that is related to carbon utilization than the taxonomic distance between the species. For some functions, clear indications of gene loss and duplication can be found. For instance, the glucuronoyl esterase encoding genes appear to have been already present in ancestral fungi as they can be found in both ascomycetes and basidiomycetes, but many species no longer have these genes (Fig. 2A) (Duranova *et al.* 2009). While many fungi have only a single copy of this gene, multiple copies were identified in some fungi (Fig. 2B) (Duranova *et al.* 2009). It would therefore introduce errors if these genes were taken along in genome wide phylogenies. However, for understanding how fungi adapt to their environment these genes are very interesting.

What are the challenges in your research area?

I see two main challenges in my research area:

- To identify the core set of genomes in fungal genomes that can be used to do genome-based phylogeny and systematics. This will provide a more in depth understanding of taxonomy and general fungal evolution.
- To identify which genes are highly susceptible to environmental pressure and therefore are likely to evolve with a higher rate than 'general evolution'. This likely does not only applies to carbon utilization, but also the utilization of other nutrients, temperature, pH, etc.

What needs do you see in the next future for your research area?

The number of fungal genomes will increase exponentially in the near future, but the ability to perform manual annotation and corrections will likely decrease. It is therefore essential that better automatic gene calling and annotation algorithms are developed. The quality of a genome will affect all subsequent analyses, so maintaining high quality without manual curation should be a main concern. In addition, platforms for comparative genomics and transrciptomics need to be further developed to deal with the ever growing datasets.

Why taxonomy is relevant to society

The relevance of taxonomy for society is mainly in being able to distinguish the different fungi that share our environment. In many cases, a pathogenic fungus can be closely related to a harmless fungus and being able to distinguish these is essential. In plant pathogenic fungi, the host range of two sister species can differ dramatically, so also there this distinction is crucial.

Taxonomy provides the basic characteristics to identify fungi and with that their potential for applications as well as their potential risk to health or the environment. In the genomic age we have already seen that classical taxonomy needs to be revisited based on genomic insights and the genomes can help in quicker identifying the traits of a fungus.

In another context, genomic taxonomy can help us understand evolution. Due to the short generation times and the rapid evolution of fungal genomes, this can is an ideal model system to study this important topic.

References

Battaglia, E., Benoit, I., Gruben, B.S. & de Vries, R.P. (2010) Plant cell wall derived sugars as substrates for fungi and industry. *Sugar Industry: Challenges, Issues and Prospects.* (ed. P.T. Jenkins), pp. 65-94. Nova Science Publishers, Inc., Hauppauge, NY.

Battaglia, E., Visser, L., Nijssen, A., van Veluw, J., Wösten, H.A.B. & de Vries, R.P. (2011) Analysis of regulation of pentose utilisation in *Aspergillus niger* reveals evolutionary adaptations in the Eurotiales. *Stud. Mycol.,* **69,** 31-38.

Coutinho, P.M., Andersen, M.R., Kolenova, K., vanKuyk, P.A., Benoit, I., Gruben, B.S., Trejo-Aguilar, B., Visser, H., van Solingen, P., Pakula, T., Seiboth, B., Battaglia, E., Aguilar-Osorio, G., de Jong, J.F., Ohm, R.A., Aguilar, M., Henrissat, B., Nielsen, J., Stalbrand, H. & de Vries, R.P. (2009) Post-genomic insights into the plant polysaccharide degradation potential of *Aspergillus nidulans* and comparison to *Aspergillus niger* and *Aspergillus oryzae*. *Fungal Genet Biol,* **46 Suppl 1,** S161-S169.

de Vries, R.P. (2003) Regulation of *Aspergillus* genes encoding plant cell wall polysaccharide degrading enzymes; relevance for industrial production. *Appl. Microbiol. Biotechnol.,* **61,** 10-20.

de Vries, R.P. & Visser, J. (2001) *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microb. Mol. Biol. Rev.,* **65,** 497-522.

Duranova, M., Spanikova, S., Wosten, H.A., Biely, P. & de Vries, R.P. (2009) Two glucuronoyl esterases of Phanerochaete chrysosporium. *Arch Microbiol,* **191,** 133-140.

Espagne, E., Lespinet, O., Malagnac, F., Da Silva, C., Jaillon, O., Porcel, B.M., Couloux, A., Ségurens, B., Poulain, J., Anthouard, V., Grossetete, S., Khalili, H., Coppin, E., Déquard-Chablat, M., Picard, M., Contamine, V., Arnaise, S., Bourdais, A., Berteaux-Lecellier, V., Gautheret, D., de Vries, R.P., Battaglia, E., Coutinho, P.M., Danchin, E.G.J., Henrissat, B., El Khoury, R., Sainsard-Chanet, A., Boivin, A., Pinan-Lucarré, B., Sellem, C.H., Debuchy, R., Wincker, P., Weissenbach, J. & Silar, P. (2008) The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biol.,* **9,** R77 (71-22).

Gruben, B.S. (2012) Novel transcriptional activators of *Aspergillus* involved in plant biomass utilization. PhD, Utrecht University.

van den Brink, J. & de Vries, R.P. (2011) Fungal enzyme sets for plant polysaccharide degradation. *Appl Microbiol Biotechnol*.

Vankuyk, P.A., Benen, J.A., Wosten, H.A., Visser, J. & de Vries, R.P. (2012) A broader role for AmyR in *Aspergillus niger*: regulation of the utilisation of D-glucose or D-galactose containing oligo- and polysaccharides. *Appl Microbiol Biotechnol,* **93,** 285-293.

The polyphasic approach in the “omics” era – up to date, or out of date?

– Peter Kämpfer

Justus-Liebig-University Giessen, Department Microbiology of Recycling-Processes, Giessen, Germany

The term polyphasic taxonomy, which was coined almost 40 years ago aimed at the integration of many levels of information (from molecular to ecological data) thereby allowing a holistic view which needs to be revisited in the light of the enormous potential of the novel information associated with large data sets from the “omics” approaches.

Theses

- Prokaryotic taxonomy (consisting of its basic elements classification, nomenclature and identification) serves for many purposes. It should be and remain stable and predictable.

- An ideal taxonomy would involve one system (a hierarchical system). A stable and accepted nomenclature should preferably also comprise the numerous “uncultured” organisms. This requires an overall accepted language (nomenclature) with a common meaning of the terms in use. The bases of this have been established in the ‘‘The Bacteriological Code of Nomenclature’’, however, names of uncultured organisms are not covered by the Code. Despite this fact, all microbiological disciplines use “names” of taxa, but the principles of nomenclature have been almost forgotten.

- The possibility to generate whole genome sequences in a very short period of time leads to a strong tendency to base the taxonomic system more and more on sequence data, which are associated with (organism and gene) names, when deposited in public databases. This may lead to major problems in the future in case of wrong or low quality sequences associated with wrong names! Efforts from microbiological societies, funding bodies, academies etc. are necessary to support taxonomic systems and or databases that are maintained and scientifically sustainable curated.

- A comprehensive understanding of all the information behind sequence data and other “omics” data is lagging far behind their accumulation. Genes and genomes may (or may not) function only in a given ‘‘environment’’, with the cell as basic entity for the display of this potential.

- Prokaryotic taxonomy should still have its focus on the whole organism (and not only on a genomes….transcriptomes…..proteomes…..). In this context, natural selection drives evolution selecting the existing phenotypes and it is the phenotype that ‘‘exhibits’’ this process both in a given cellular and also environmental context. The importance of the phenotype (which is not only restricted to a physiological/ biochemical fingerprint of an organisms) needs to be stressed again.

What are the current main challenges in your research area?

The introduction of 16S rRNA sequence into the prokaryotic taxonomy as an outstanding molecular marker, has served now for more than 20 years as a “backbone” of prokaryote classification and in the scientific community it is widely accepted, that all prokaryotes are classified into the domains “Archaea” or “Bacteria”, on the basis of the sequence comparisons of this molecule. These domains can then be subdivided in a hierarchical manner into the descending taxonomic ranks: “phylum”, “class”, “order”, “family”, “genus”, “species” and some of these ranks are further (not consistently) subdivided into lower ranks using the suffix “sub-“ like: e.g. “suborder”, or “subspecies”. It is essential to note, that all taxonomic ranks mentioned above are nothing else than categories, i.e. they are more or less subjective concepts without any foundation in fact. In establishing a hierarchical (or any other) system, described by taxonomic ranks, it is often taken as “self-evident” that an overall accepted nomenclature with a common meaning of the terms for those taxonomic ranks is in use. What is sometimes not explicitly expressed and/or perhaps not appreciated is the fact, that all microbiological disciplines are taking advantage of this overall agreed nomenclature. Even in case of a sequence, which may or may not be submitted to a database, it is the rule rather than an exception that a name of a taxonomic rank is associated with this sequence.

Nomenclature plays an important role, and without an overall agreed “language”, the whole communication across the microbiological disciplines would not be possible. To ensure a stability and predictability, the nomenclature of the “lower” ranks i.e. “family”, “genus”, “species” and “subspecies” is regulated by a rules of the nomenclatural system, “The Bacteriological Code of Nomenclature”, that has been developed several decades ago and still serves all microbiologists as a solid and indispensable foundation for a common scientific exchange of data and information. This has to be maintained and it is essential that databases that are used are correct in regard to the entries and their links to nomenclature.

With the development of new sequencing methods, it has now become feasible to look in detail at the genome sequence and to generate gene and genome sequences in a relatively short period of time and it can be foreseen that the wealth of the new data can and will be used for a critical evaluation of the taxonomic system. Again databases that are used must be curated and it is essential for the future that the entries are correct.

The more genome sequences become available, the more difficulties in comparing whole genomes become also obvious. Only a limited number of genes occur in all genomes. The more genome sequences are available, the smaller the number of genes, which are present in all genomes. Information content for phylogenetic analyses is often not clear. Recognition of paralogous genes is a major problem and conflicting tree topologies may lead to false conclusions. Other open questions are: What genes belong to the conserved genome core and to the accessory dispensible genetic elements? In particular, the impact of processes like lateral gene transfer, gene duplication, recombination, rearrangements of genes in the genome may lead to intrinsic difficulties in genome sequence comparisons. It should always be stressed, that prokaryotic taxonomy serves for many purposes. It should be stable and predictable. A most comprehensive phenotypic and genotypic characterization (in the framework of a polyphasic approach is still necessary in characterization and classification (which is a prerequisite of identification). The introduction of quality standards (at all levels) are highly desirable. Genomic (and other omic) approaches will provide a rich source of data that should be carefully investigated in regard to the information behind. It is very difficult to predict higher levels of

„information“ from genome data sets. Molecules may tell us nothing about cells and their behaviour. In essence, the conversion of data into knowledge (at different levels) constitutes a great challenge for future biological research, also for taxonomy.

What needs do you see in the next future for your research area?

A most comprehensive phenotypic and genotypic characterization (in the framework of a polyphasic approach) is still necessary in characterization and classification (which is a prerequisite of identification). The introduction of quality standards are highly desirable. Again, genomic (and other omic) approaches will provide a rich source of data that should be carefully investigated in regard to the information behind. It is very difficult to predict higher levels of „information“ from genome data sets. This is a (general) major task for research in the future.

Please describe in few words why taxonomy is relevant to society.

The term taxonomy is often used synonymously with systematics, but it should be regarded more as a specific part of the latter and comprises the orderly arrangements of (defined) units in addition to the nomenclature, i.e. labelling of these units defined by classification, and also identification of these units defined by classification and labeled by nomenclature. As already mentioned above, it is sometimes not explicitly expressed and/or perhaps not appreciated, that all microbiological disciplines are taking advantage of this overall agreed nomenclature. Without this fact an unambiguous communication would not be possible, neither in the scientific area nor anywhere else.

Entwicklung und Biotechnologie von Pilzen - Fungal development, fungal biotechnology

– Ulrich Kück

Ruhr-Universität Bochum, Fakultät für Biologie und Biotechnologie, Lehrstuhl für Allgemeine und Molekulare Botanik, Bochum

Einführung

Nur ein geringer Teil der auf der Erde lebenden Arten ist bisher erfasst. Es wird geschätzt, dass ca. 90% der auf der Erde existierenden Arten bisher noch unentdeckt ist. Die Tabelle 1 macht deutlich, dass der Großteil der Wirbeltiere bekannt ist, dass jedoch bei den Mikroorganismen (Bakterien, Archaeen, Pilze und Algen) die „Dunkelziffer“ noch recht groß ist (Bull et al. 1992). Dies liegt unter anderem an der oft schwierigen Kultivierbarkeit mikrobieller Lebewesen. Bemerkenswert ist, dass die Pilze innerhalb der mikrobiellen Organismen scheinbar die größte Gruppe darstellen. Zurzeit sind 70.000 bis 140.000 Pilzarten bestimmt, konservative Schätzungen gehen jedoch davon aus, dass auf der Erde ca. 1,5 Mio. Pilzspezies leben. In anderen Publikationen wird sogar eine Zahl von 9,9 Mio. Pilzarten geschätzt (Hawksworth 2001). Pilze spielen nicht nur als Saprophyten eine wesentliche Rolle im ökologischen Haushalt der Natur, sondern haben auch für die Anwendung eine übergeordnete und entscheidende Rolle, z. B. im Agrarbereich, wo sie als Mykorrhiza-Symbionten wesentlich zum Ertrag von Nutzpflanzen beitragen. Darüber hinaus besitzen Pilze in der Biotechnologie eine herausragende Bedeutung. Sie werden z. B. in der Nahrungsindustrie (Brotbacken, Käseproduktion), bei der Herstellung alkoholischer Getränke, bei der Herstellung organischer Säuren (Zitronensäure für die Getränke- und Lebensmittelindustrie) und bei der Fermentation von Nahrung (Sojasauce) genutzt. In der pharmazeutischen Industrie gibt es eine Vielzahl verschiedener Anwendungen, um z. B. Antibiotika, Alkaloide, Immunosuppressiva, Steroide, Pflanzenwachstumsregulatoren oder Statine zu produzieren (Kück et al. 2009). Für die Anwendung ist die zweifelsfreie Bestimmung und Klassifizierung von Pilzen von entscheidender Bedeutung. Nur so können nicht nur reproduzierbare Produktionsverfahren gewährleistet werden, sondern es ist auch nur so möglich, durch Naturisolate neue Naturstoffe zu entdecken, die für die Anwendung von Bedeutung sind.

Traditionell wurde bei den Pilzen in der Taxonomie der Fruchtkörper (Teil des Sexualzyklus) für die taxonomische Einteilung genutzt, und ein sehr allgemeiner phylogenetischer Stammbaum auf dieser Basis ist in der Abb. 1 dargestellt. Allerdings vermehren sich viele Pilze ausschließlich asexuell durch Konidiosporen, sodass vor allen Dingen die Konidienträger, die Produzenten der Konidiosporen, für die taxonomische Klassifizierung herangezogen wurden. Die molekulare Phylogenie, die in den neunziger Jahren eine Vielzahl von neuen DNA-Sequenzen lieferte, führte anfänglich zu großer Verwirrung und Unsicherheit bei der Klassifikation von Pilzen. Allerdings sind inzwischen diese Daten weiter gefestigt worden, sodass heute mindestens 14.054 verschiedene Pilzspezies in der Genbank hinterlegt sind. Dies entspricht in etwa 13% der im Augenblick bekannten Arten (Hawksworth 2006).

Eine besondere Schwierigkeit ergab sich insbesondere dadurch, dass bei vielen Pilzen zwei verschiedene Formen, die als unterschiedliche Spezies gehalten wurden, bekannt sind. So gibt es teleomorphe Stadien, die sich sexuell vermehren können, sowie anamorphe Stadien, bei denen ausschließlich asexuelle Fortpflanzung auftritt. Für viele Ascomyceten, die im Volksmund als „Schimmelpilze“ bezeichnet werden, sind nur die anamorphen Stadien bekannt, ein Sexualzyklus ist in vielen Fällen nicht beschrieben worden. Jedoch nimmt die Zahl der Fälle zu, wo anamorphe Ascomyceten, von denen davon ausgegangen wurde, dass sie ausschließlich durch Konidiosporen vermehrbar sind, auch einen Sexualzyklus besitzen (Kück und Pöggeler 2009). Dies hatte zur Folge, dass aufgrund der dualen No-

menklatur die asexuellen und sexuellen Stadien unterschiedliche lateinische Namen trugen. Normalerweise wird der teleomorphe Name dem des anamorphen vorgezogen.

Tabelle 1: Geschätzte Zahl und Anteil der bereits bestimmten Arten innerhalb verschiedener organismischer Gruppen (verändert nach Burton et al. 2002)

Group of organisms	Estimated species	Accessible (known) species (as % of total)
Animals (mammals, birds, fishes)	3.5×10^4	>90
Arthropods/invertebrates	$10^6 - 10^7$	10
Nematodes	5×10^5	3
Higher plants	2.7×10^5	>90
Algae[3]	$10^4 - 10^5$	[70]
Bryophytes	2.5×10^4	70
Fungi[3]	1.5×10^6	[5]
Bacteria[3]	$10^4 - 10^5$	[1-10%]
Archaea[3]	$10^5 - 10^6$	[0.1-1%]
Viruses[3]	$10^5 - 10^6$	[4]

Was sind die aktuellen Herausforderungen in unserem Forschungsfeld?

Am Beispiel des industriellen Penicillin-Produzenten *Penicillium chrysogenum* soll die Schwierigkeit bei der taxonomischen Einordnung von anamorphen und teleomorphen Formen dokumentiert werden.

Vertreter der Gattung Penicillium zählen umgangssprachlich zu den Schimmelpilzen und wurden bereits zu Beginn des 19. Jahrhunderts wissenschaftlich beschrieben (zitiert in Pitt 1979). Die Gattung Penicillium ist durch die typischen Konidienträger (Konidiophoren) gekennzeichnet, die die Form eines Pinsels besitzen, der im ausgeprägtesten Fall in sogenannte Rami, Metulae und Phialiden gegliedert ist. Später hat der Botaniker Brefeld in der 1874 verfassten Schrift „Die Entwicklungsgeschichte von Penicillium“ die ersten sexuellen, also teleomorphen Stadien von Penicillium-Spezies beschrieben (Brefeld 1874). Hierzu zählte *Penicillium chrystaceum*, der später umbenannt wurde in *Eupenicillium chrystaceum*. Viele Penicillium-Arten haben als Saprophyten, ähnlich wie Arten der Gattung Aspergillus eine überragende Rolle bei der Umsetzung organischen Materials. Sie sind ein wichtiger ökologischer Baustein im Stoffkreislauf der belebten Natur. Viele Penicillium-Arten haben darüber hinaus eine Bedeutung für die Anwendung. Beispielhaft sollen *Penicillium roqueforti* und *Penicilliium camemberti* genannt werden, die nicht nur für den Geschmack, sondern auch für die Fermentation von Milchprodukten bei der Käseherstellung verantwortlich sind. Ein anderes Beispiel ist *Penicillium purpurogenum*, der biotechnologisch eine Bedeutung als Xylanaseproduzent besitzt. Schließlich werden mehrere Penicillium-Arten kommerziell als Pharmaproduzenten genutzt. *Penicillium brevicompactum* z. B. produziert Mykophenolsäure, welche als Immunosuppressivum eingesetzt wird. Eine mindestens ebenso große Bedeutung besitzt *Penicillium citrinum*, der Produzent eines Statins, das besonders in der westlichen Welt als Cholesterinsenker genutzt wird. Die historisch bedeutendste Leistung liefert *Penicillium chrysogenum*, der das β-Laktam-Antibiotikum Penicillin produziert. Mit einem Weltmarkt von ca. 13,5 Mrd. Euro haben Penicilline eine überragende Bedeutung im Antiinfektiva-Markt (Barber et al. 2004). Für *Penicillium chrysogenum* sind verschiedenste molekulare Techniken entwickelt worden, um diesen Pilz gentechnisch zu verändern. Dies ist bei der Ent-

wicklung von Produktionsstämmen wichtig, die sowohl durch traditionelle Verfahren (Zufallsmutagenese), aber auch gezielte genetische Manipulation optimiert werden. Wir konnten vor kurzem sogenannte Kreuzungstyploci identifizieren, die „Geschlechtsgene" tragen (Hoff et al. 2008). Diese erhalten in der Regel die Bezeichnung MAT1-1 und MAT1-2. Grundsätzlich tragen die einzelnen Isolate nur eines dieser beiden Geschlechtsgene. Für den Sexualzyklus müssen deshalb Isolate, die unterschiedliche Kreuzungstyploci tragen, miteinander fusionieren, um entsprechend den Sexualzyklus einzuleiten. Ein derartiges Entwicklungssystem mit zwei Kreuzungstypstämmen wird als heterothallisch bezeichnet. Durch die gezielte Sequenzierung von verschiedenen Isolaten von *P. chrysogenum, P. citrinum* und *P. brevicompactum* konnten wir zeigen, dass alle drei Arten heterothallisch sind und sich durch spezifische mating-type Gene unterscheiden lassen. Von Untersuchungen bei verschiedenen pilzlichen Modellsystemen weiß man, dass die mating-type Gene für Transkriptionsfaktoren kodieren, welche Komponenten eines Sexualzyklus steuern. Zu diesen Komponenten gehören z. B. Pheromone und Pheromonrezeptoren, die für die Zell-Zell-Erkennung wichtig sind.
Das Vorhandensein von mating-type Genen legte nahe, dass z. B. *Penicillium chrysogenum* einen Sexualzyklus besitzt. Durch gezielte Anzucht und entsprechende physiologische Versuchsbedingungen gelang es tatsächlich, bei *P. chrysogenum* den Sexualzyklus einzuleiten und den Nachweis zu liefern, dass durch den Zyklus rekombinante Stämme entstehen. Damit ist nachgewiesen, dass *P. chrysogenum* neben der asexuellen Vermehrung auch einen Sexualzyklus durchlaufen kann.
Die Tatsache jedoch, dass *Penicillium chrysogenum* einen der wesentlichen industriellen Pilze darstellt, der seit langem beschrieben worden ist, lässt die Forderung zu, dass *P. chrysogenum* eine Umbenennung erfährt. Anders als bei teleomorphen Formen, die phylogenetisch nahe zu *Penicillium chrysogenum* stehen, wie die Gattungen Talaromyces oder Eupenicillium, sollte hier der ursprüngliche Artname beibehalten werden. Dies deckt sich voll und ganz mit kürzlich erhobenen Forderungen, dass nur ein einzelner wissenschaftlicher Name für Pilzarten erhalten werden sollte (Rossman und Samuels 2005). Auf dem Botanischen Kongress in Melbourne 2011 wurde die Taxonomie von Pilzen noch einmal eingehend diskutiert, danach soll ab 1. Januar 2013 nur noch ein Name je Art üblich sein. Eine ausführliche Diskussion zu diesem Thema findet sich bei Hawksworth (2011).

Welche Notwendigkeiten sind in der Zukunft in dem Forschungsgebiet erkennbar?
Um eine taxonomisch einwandfreie Beschreibung zu erzielen, sind Genomdaten in großem Maßstab notwendig. Allerdings ist diese Umsetzung nicht immer einfach zu vollziehen, zumal oft nur geringe Probenmaterialien zur Verfügung stehen, die eine genomweite Sequenzierung nicht in jedem Fall zulassen. Außerdem erlaubt die Praxis in vielen Fällen keine zeitaufwendigen Bestimmungen. Deshalb sind die mating-type Gene eine Alternative für eine schnelle Identifizierung von Arten, um die Gattungs- und auch die Artbestimmung vorzunehmen. Die konservierte Primärstruktur der Kreuzungstyploci innerhalb einer Gattung erlaubt es, einwandfreie Artbestimmungen vorzunehmen.

Warum ist die Taxonomie wesentlich für gesellschaftliche Ansprüche?
Die angewandte Biologie, hier eingeschlossen ist die Biotechnologie, die Agrarwissenschaft, die pharmazeutische Industrie oder auch die Lebensmittelindustrie, ist darauf angewiesen, einwandfrei taxonomisch bestimmte Organismen in ihren Produktionsprozessen einzusetzen. Dadurch kann nicht nur die Qualität der Produkte gewährleistet werden, sondern auch ein Produktionsverfahren, das in einem ökonomisch vertretbaren Umfang erfolgt. Im gleichen Maße sind die Grundsätze des GMPs („Good Manufacturing Practice") zu beachten, die bei der Produktion von Arzneimitteln und Wirkstoffen die Qualitätssicherung der Produktionsabläufe sichern (Patel und Chotai 2008).

Die Tatsache, dass viele Verfahren im 100.000 Liter-Maßstab erfolgen, zeigt, dass hierbei hochkomplexe und ökonomisch sehr aufwändige Verfahren durchlaufen werden, die nur unter Einsatz von einwandfrei beschriebenen Stämmen durchführbar sind.
Die Naturstoffchemie wird in Zukunft deutlich davon profitieren, dass viele Mikropilze noch gar nicht erfasst sind (Hawksworth 2004). Die Neuentdeckung und Weiterentwicklung von Naturstoffen macht es notwendig, neue Naturisolate zu gewinnen, die in den entsprechenden Testverfahren untersucht und taxonomisch bestimmt werden. Nur dadurch kann sichergestellt werden, dass auch in Zukunft mit neuen Wirkstoffen der technologische Fortschritt im medizinischen, pharmazeutischen, agronomischen, lebensmittel- und umwelttechnologischen Bereich gewährleistet wird, der Voraussetzung für eine zukunftsorientiere Entwicklung der Gesellschaft ist.

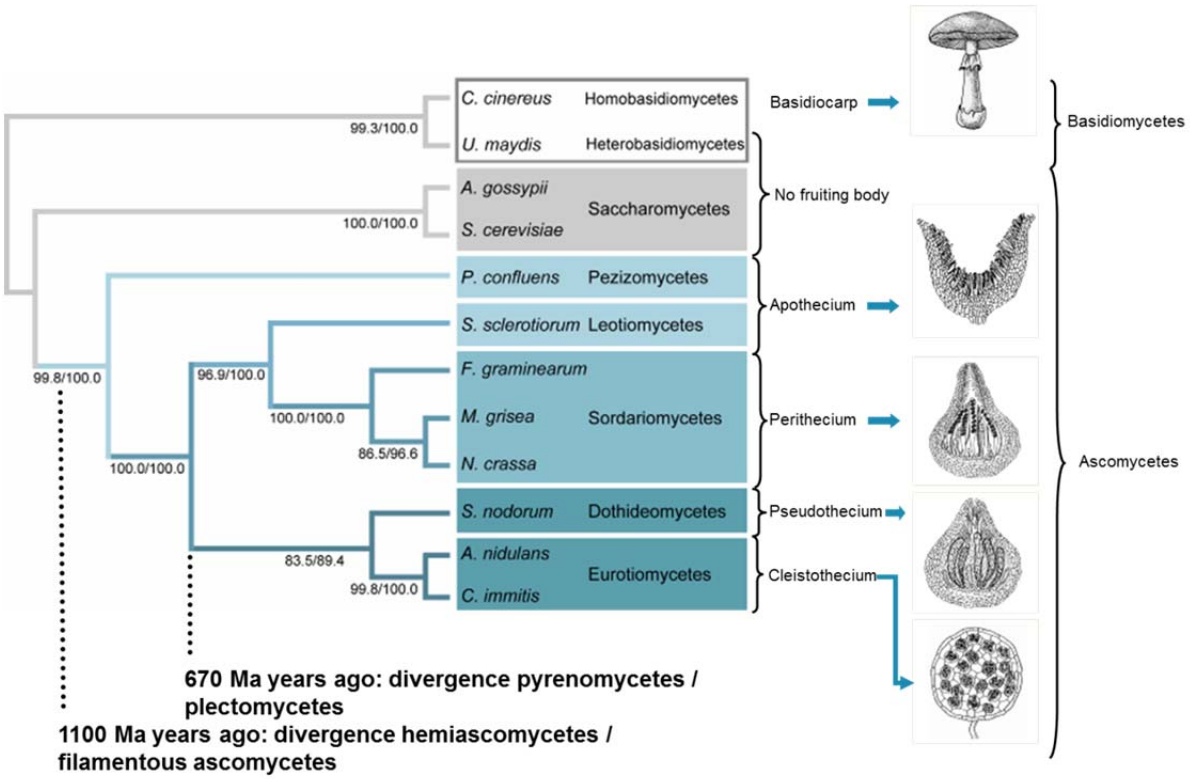

Abbildung 1: Phylogenie der Pilze, die traditionell aufgrund Ihrer Fruchtkörper taxonomisch eingeteilt wurden. Beispielhaft sind vier Fruchtkörper von filamentösen Ascomyceten (Schlauchpilze) dargestellt. Verändert nach Heckman et al. 2001, Nowrousian und Kück 2006, Pöggeler et al. 2006.

Referenzen

Barber MS, Giesecke U, Reichert A, Minas W (2004) Industrial enzymatic production of cephalosporin-based beta-lactams. Adv Biochem Eng Biotechnol 88: 179-215

Bull, AT, Goodfellow M, Slater JH (1992) Biodiversity as a source of innovation in biotechnology. Annu Rev Microbiol 46: 219-252

Brefeld O (1874) Botanische Untersuchungen über Schimmelpilze. Heft 2 “Die Entwicklungsgeschichte von *Penicillium*” Leipzig 98 pp

Burton SG, Cowan DA, Woodley JM (2002) The search for the ideal biocatalyst. Nature Biotechnology 20: 37-44

Hawksworth DL (2011) A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. IMA MycoKeys 1: 7–20

Hawksworth DL (2006) Pandora's mycological box: Molecular sequences vs. morphology in understanding fungal relationships and biodiversity. Rev Iberoam Micol 23: 127-133

Hawksworth DL (2004) Fungal diversity and its implications for genetic resource collections. Stud Mycol 50: 9-18

Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycol Res 105: 1422-1432

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB (2001) Molecular evidence for the early colonization of land by fungi and plants. Science 293: 1129-1133

Hoff B, Pöggeler S, Kück U (2008) Eighty years after its discovery Fleming's Penicillium strain discloses the secret of its sex. Eukaryot Cell 7: 465-470

Kück U, Pöggeler S (2009) Cryptic Sex in Fungi. Fungal Biol Rev 23: 55-90

U. Kück, M. Nowrousian, B. Hoff, I. Engh: Schimmelpilze. Springer Verlag, Heidelberg (2009)

Link HF (1809) Observations in Ordines plantarum naturales, Dissertatio 1ma, Berlin Ges Nat Kde 3: 1-42

Nowrousian M, Kück U (2006) Comparative gene expression analyses of fruiting body development in two filamentous fungi. FEMS Microbiol Lett 257: 328-335

Rossman AY, Samuels GJ (2005) Towards a single scientific name for species of fungi. Inoculum 56: 3-6

Patel KT, Chotai NP (2008) Pharmaceutical GMP: past, present, and future--a review. Pharmazie 63:251-255

Pitt JI (1979) The genus Penicillium and its teleomorphic states *Eupenicillium* and *Talaromyces.* Academics Press London

Pöggeler S, Nowrousian M, Kück U (2006) Fruiting body development in ascomycetes. In: Kües U, Fischer R, THE MYCOTA I (2nd ed.) Springer Verlag, Berlin Heidelberg, pp 325-355

Typing and identification of microorganisms at sub-species levels: considerations for biodiversity, biotechnological, clinical and epidemiological studies

– Edward Moore

University of Gothenburg, Institute of Biomedicine, Göteborg, Sweden

The basis of the systematic and taxonomic organisation of microorganisms is derived from estimates of their phylogenetic relationships. In turn, the phylogenetic relationships of microorganisms are based upon genotypic and, ultimately, genomic characterisations. In the current age of 'Next-Generation Sequencing' technology, we will certainly see the criteria of genome sequence determinations and comparisons included in new microbial descriptions and the validations of new species names. However, the application of proteomic biomarkers for analyses of microorganisms offers rapid, inexpensive and high-resolution analyses of microorganisms, at the species- and sub-species levels. For the analyses of microorganisms in biodiversity, biotechnological, clinical and epidemiological studies, reliable characterisations, at sub-species levels, is increasingly essential. Such analyses are absolutely dependent upon the effective construction of comprehensive genomic data.

Theses

- The typing and identification of microorganisms at sub-species levels is increasingly essential.
- Proteomic analyses offer rapid, inexpensive and high-resolution characterisation of microorganisms, at sub-species levels.
- Growing genomic databases enable new applications for exploiting proteomics approaches for reliable, rapid and inexpensive analyses of microorganisms.

Diversification in nonpathogenic bacteria - implications for taxonomy, bioinformatics and data repositories

– Jörg Overmann

Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

Phylogenetic comparisons of 16S rRNA gene sequences show that bacterial lineages cluster in a hierarchical fashion. Multilocus sequence analysis (MLSA; Maiden 2006) of housekeeping genes (such as *rpoB*, *recA* or *atpD*) provides a refined insight into the evolution of bacterial lineages since these housekeeping genes cover a larger fraction of the genome, exhibit higher rates of sequence change and hence provide a higher resolution for sequence comparisons. In addition, the simultaneous analysis of concatenated sequences of multiple housekeeping genes is less affected by lateral gene transfer. So far, MLSA has mostly been applied to pathogenic bacteria like *Burkholderia pseudomallei*, *Haemophilus influenzae*, *Helicobacter pylori*, *Neisseria meningitidis*, and various *Salmonella*, *Staphylococcus* and *Streptococcus* species (Maiden 2006; Kilian *et al*., 2008; Keymer & Boehm, 2011; Achtman, 2008; Nübel *et al*., 2010; compare also the online MLSA databases www.mlst.net and pubmlst.org). Various models of bacterial evolution have been developed to explain the hierarchical clustering of bacterial lineages detected by MLSA (Fig. 1) (Gevers et al., 2005).

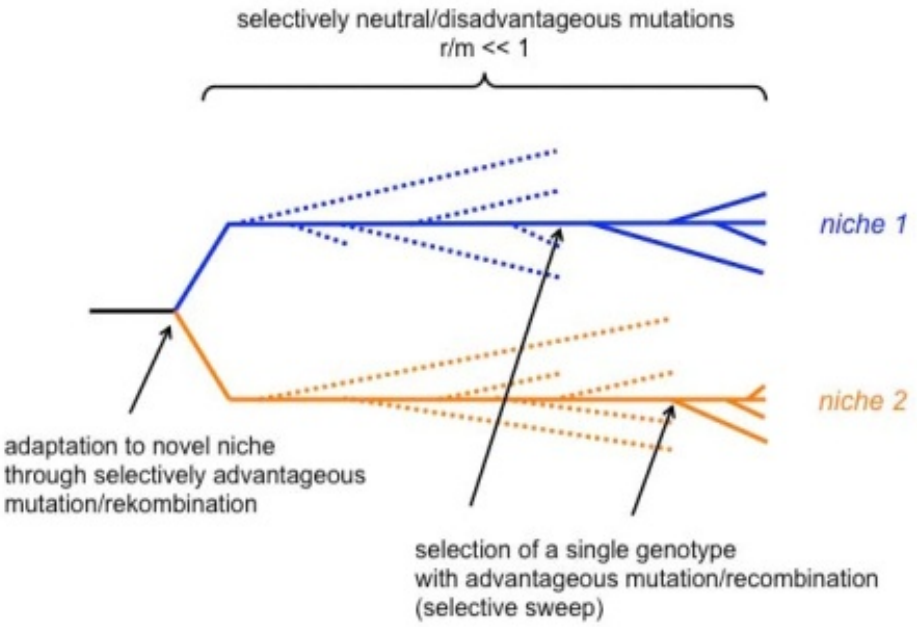

Figure 1. One exemplary model ("ecotype model") for the diversification of bacteria under different evolutionary mechanisms. The phylogenetic tree is based on the sequence divergence of housekeeping genes. Two major lineages (colored blue and orange) occupy different niches and survive as discrete lineages as long as recombination rates are not so large that niche-specifying genes are exchanged to a substantial amount between lineages. In this scenario, the evolutionary force is natural selection which keeps the individual lineages cohesive. Within lineages, individual genotypes occasionally gain a selective advantage but do not recombine rapidly with less advantageous genotypes. As a result, the latter (depicted as dashed lines) become extinct during selective sweeps.

In some instances, the results of MLSA are not congruent with bacterial taxonomy, suggesting that some established bacterial taxa (in particular named species) do not correspond to evolutionary cohesive entities. In fact, the ratio of recombination to mutation rates (*r*/*m*; Fig. 1) varies widely between bacteria and ranges from a minimum value of 0.02 in *Staphylococcus aureus* to > 63 in *Flavobacterium psychrophilum* and *Pelagibacter ubique* (Vos and Didelot, 2009). Interestingly, recombination seems to be more significant for the evolution of aquatic than of terrestrial bacteria. In addition, evolutionary forces such as genetic drift may also govern the evolution of bacterial lineages, and hence may also act as forces that keep lineages cohesive and separated from other lineages. Thus, evolutionary cohesive entitities may exist at different phylogenetic levels and driven by different evolutionary forces. As a consequence, such cohesive entities are likely not to be in congruence with taxa, e.g., named bacterial species.

From the above, the following key questions for the future development of bacterial taxonomy can be deduced:

- To which extent are different evolutionary entitities, from the onset of their origin, kept cohesive by natural selection, recombination, or genetic drift and to which extent are they genetically isolated?
- Can the cohesion of evolutionary entitities be traced back to their ecology and how can specific adaptive traits be determined?
- To what extent are evolutionary entities congruent with taxonomic categories of named bacterial species or even higher taxa?
- Finally, will it be possible to develop a concept for bacterial taxonomy that allows to replace pragmatic but rigid and artificial cutoffs based on discrete numerical values by evolutionary cohesive entities?

Recent analyses by this laboratory of natural populations of aquatic *Sphingomonadaceae* suggest that diversification occurs within a single bacterial species and even within the same 16S rRNA gene sequence type. 95 *Sphingomonadaceae*strains were recovered from two freshwater lakes and subjected to high-resolution MLSA of nine housekeeping genes and a parallel phenotypic characterization. 52 strains had identical 16S rRNA gene sequences but based on MLSA formed three genetically distinct subpopulations that were separated due to low rates of genetic recombination. Discrete seasonal abundance patterns of different members of the population indicate that phenotypic differences exist. However, no consistent physiological differences were detected between the three subpopulations despite extensive physiological testing. Whereas distinct phenotypic differences of subpopulations have been shown for other freshwater bacteria (Jaspers and Overmann 2004), alternative phenotypic traits need to be postulated that provide the selective advantage in the case of the three individual subpopulations of *Sphingomonadaceae*. These examples demonstrate that incipient speciation occurs among the same 16S rRNA phylotype, and result in cohesive, genetically largely separated, and sympatric subpopulations. These subpopulations that exist within traditional bacterial species and even within one 16S rRNA sequence type need to be considered in bacterial taxonomy. These MLSA-based studies support the view that criteria based on cohesion may be more adequate for delineating evolutionary entities (hence taxa) than simple and rigid cutoffs based on discrete numerical values of sequence similarity.

Yet, even MLSA of a larger number of up to 10 housekeeping genes still covers only < 1% of a bacterial genome and therefore can only provide a very limited insight into bacterial evolution. State-of-the-art omics approaches now offer the potential to analyse bacterial evolution on a much broader scale, to generate hypotheses regarding the advent and persistence of evolutionary units, and ultimately to arrive at a bacterial taxonomy that is based on the evolutionary history - i.e. a natural system of bacteria.

The potential insights into bacterial evolution that may be provided by omics approaches can be exemplified using the understudied group of green sulfur bacteria. Green sulfur bacteria (*Chlorobiaceae*) depend on reduced sulfur compounds (in one case Fe^{2+}) as photosynthetic electron donor, CO_2 as carbon source and light as energy source (Overmann 2001a). *Chlorobiaceae* typically occur where light reaches sulfide-containing anoxic waters or sediments and supposedly occupy a very narrow ecological niche. Each strain shares 50 – 90% of its total gene content with those of other *Chlorobiaceae*. Apparently, these bacteria thus share a large number of biochemical and physiological traits. Though different isolates of *Chlorobiaceae* show distinct preferences with regard to light intensity, salinity, sulfide concentration or a pelagic versus benthic mode of life, it remains obscure how the phylogenetic diversity (of an estimated 3.500 different 16S rRNA gene sequence types) can be maintained. Genome analyses yielded first indications for one particular mechanism of niche adaptation in *Chlorobiaceae* (Wenter et al. 2010). A representative of the green sulfur bacteria that forms multicellular associations (so-called phototrophic consortia; Overmann, 2001b) with a motile betaproteobacterium harbors large and unique "symbiosis" genes that show similarity to bacterial virulence factors (RTX-toxins and hemagglutinins) and seem to be involved in the cell-cell-interaction with the chemoheterotrophic partner. These genes appear to have been laterally acquired from proteobacteria and allowed the green sulfur bacterium to gain entirely novel functions and to occupy a novel ecological niche. Obviously, genetic modules that so far were thought to be restricted to pathogens of eukaryotes are employed by green sulfur bacteria in a different context for beneficial interactions with another bacterial species. Similar to the genomes of other bacteria, up to 25% of the genes of *Chlorobiaceae* could not be assigned to known clusters of orthologous genes (Bryant et al. 2012). These so far unidentified genes most likely comprise additional niche-specific genes.

Understanding the formation and breadth of evolving entities among bacteria has major practical implications, in particular regarding the emergence of novel pathogens, the patenting of bacterial strains used in biotechnological processes, or with respect to legal aspects of biosafety or biosecurity. Yet, omics-supported taxonomy currently faces several main challenges:

1. In order to identify cohesive groups of bacteria, larger numbers of genomes of closely related strains must become available. Novel cultivation approaches yielding a higher cultivation success for aquatic and soil samples should be employed but also need to be further improved. Given the recent progress in single-cell genomics, such cultivation-based approaches can be complemented by multiple displacement amplification (MDA) of individual genomes from natural samples. MDA-generated genome sequences would be subject to other biases than cultivation approaches.
2. Bacterial ressource centers need to build up the capacity to collect and provide representative sets of isolates from bacterial populations. This will encompass the development of improved cultivation, guided by the results of meta-omics studies of the respective environments, as well as cryoconservation techniques.
3. Second and third generation sequencing technology already provides a significantly increased capacity to generate raw sequence data that is sufficient to analyse populations based on the entire genome sequences of their members (population genomics). However, bioinformatic capacities in many institutions are still very limited and cannot cover the needs for rapid assembly and extensive comparison of the required number of bacterial genomes. Hence, a sustained development of bioinformatic expertise most likely will be the key step when establishing an omics-supported taxonomy.
4. Bioinformatic pipelines for the rapid analysis of evolutionary patterns of larger numbers (≥ 100) bacterial genomes need to be developed urgently.
5. High throughput methods for testing hypotheses on ecophysiological phenotypes, as derived from physicochemical environmental data, need to be developed.

6. The driving forces of bacterial evolution can only be fully understood if the metadata associated with each given strain/genome (physicochemical environmental data, ecophysiological phenotypic data, biogeography) are collected, merged and made available through dedicated databases. A massive mobilization, curation and assembly of taxon-associated data will be instrumental to enable next generation biodiversity informatics and improve bacterial taxonomy, and to distribute taxonomic knowledge.

References

Achtman M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Ann Rev Microbiol* **62**:53-70.

Bryant, D. A., Liu, Z., Li, T., Zhao, F., Garcia Costas, A. M., Klatt, C. G., Ward, D. M., Frigaard, N.-U., and Overmann, J. (2012) Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria.* In: *Advances in Photosynthesis and Respiration*, Vol. 33, *Functional Genomics and Evolution of Photosynthetic Systems*, (Burnap, R. L. and Vermaas, W.F.J., eds.), pp. 47-102, Springer, Dordrecht, The Netherlands.

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ *et al.*(2005). Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**:733-739.

Jaspers, E., Overmann, J. (2004) The ecological significance of "microdiversity": identical 16S rRNA gene sequences represent bacteria with highly divergent genomes and physiology. Appl. Environ. Microbiol. 70, 4831-4839

Keymer DP, Boehm AB. (2011). Recombination shapes the structure of an environmental *Vibrio cholerae* population. *Appl Environ Microbiol* **77**:537-544.

Kilian M, Poulsen K, Blomqvist T, Håvarstein LS, Bek-Thomsen M, Tettelin H *et al.* (2008). Evolution of *Streptococcus pneumoniae* and its close commensal relatives. *PLoS ONE* **3**:e2683.

Maiden MCJ. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**:561-588.

Nübel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK *et al.* (2010). A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog* **6**:e1000855.

Overmann, J. (2001a) Green Sulfur Bacteria. In: Boone, D.R., Castenholz, R.W. (eds) Bergey´s Manual of Systematic Bacteriology, 2nd ed., Vol. 1. Williams and Wilkins, Baltimore, pp. 601-630

Overmann, J. (2001b) Phototrophic consortia: A tight cooperation between non-related eubacteria. In: Seckbach, J. (ed.) Symbiosis: Mechanisms and Model systems. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 239-255

Vos M, Didelot X. (2008). A comparison of homologous recombination rates in bacteria and archaea. *ISME J* **3**:199-208.

Wenter, R., Hütz,K., Dibbern, D., Reisinger, V., Li, T., Plöscher, M., Eichacker, L., Eddie, B., Hanson, T., Bryant, D., Overmann, J. (2010) Expression-based identification of genetic determinants of the bacterial symbiosis in `*Chlorochromatium aggregatum*´. Environ. Microbiol. 12: 2259-2276

Past, presence and future of microbial taxonomy

– Ramon Rosselló-Móra

Mediterranean Institute for Advanced Studies (IMEDEA), Marine Microbiology Group, Esporles, Spain

The taxonomy of prokaryotes is one of the youngest taxonomic disciplines among the biological sciences. Despite the existence of a bacterial world was clear already in the XIXth century, serious classifications started in the middle of the XXth century due to the technological developments that enhanced both culturing bacteria in the laboratory and exhaustive studies of their phenotype and, in some extend, genotype. Actually, the first classifications were mostly based on the overall similarities of the metabolic and morphologic traits of the isolates. However, the most important developments in circumscribing species occurred with the establishment of a series of techniques directed to the comparison of genomes just after the discovery of the DNA as a molecule with genetic information content. Since the late 1960's prokaryotic species were circumscribed based on the study of as many physiological and morphological traits as possible to observe discriminative phenotypes. In addition, the techniques of hybridization between DNAs helped in grouping several isolates within a single unit based on their genome similarities. In this regard, there was a good correlation between a phenotypic coherence among the strains in study and a DNA-DNA hybridization (DDH) similarity of about 70%. These correlations prompted taxonomists to recommend the use of this value as a putative genomic boundary (Wayne et al., 1987). The use of DDH to circumscribe species became a gold standard for any new classification of species based on more than a single specimen, and the 70% boundary became as well a numerical threshold commonly used. The application of such numerical threshold in taxonomy of prokaryotes has had a paramount influence in the way species had been described equivalent to the sexual isolation criterion for plants and animals (Rosselló-Móra, 2012).

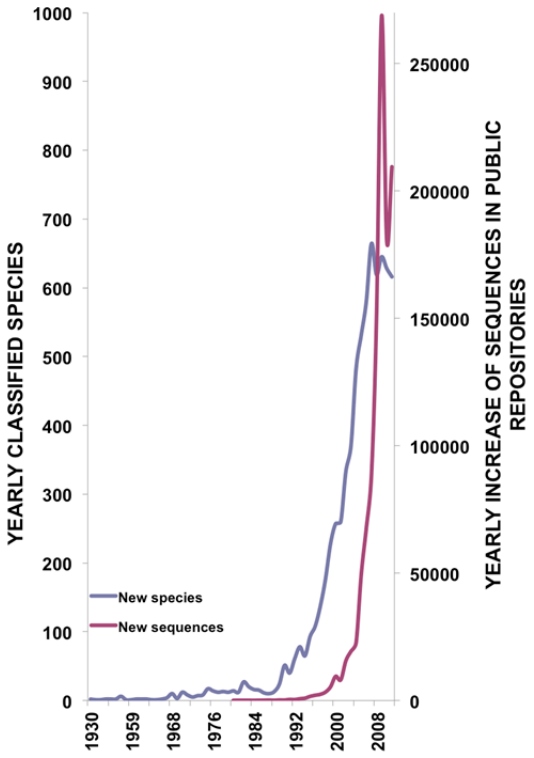

Figure 1: yearly increase of new species descriptions (blue line, left axis) and new deposited 16SrRNA sequences in public epositories (red line, right axis).

A second brakethrough in the taxonomy of prokaryotes occurred with the recognition that the sequence of the 16S rRNA gene could be used as phylogenetic marker that could reconstruct the genealogies of the organisms harbouring it (Woese et al., 1987). The application of phylogenetic reconstructions in taxonomic descriptions started to become a routine as it could place the new species in a genealogical frame based on objective genetic comparisons. This fact enhanced importantly the activities of classification of new species, and since the early 1990's the rate of new classifications increased aritmetically (Figure 1). However, the use of 16S rRNA gene sequences was not only restricted to taxonomic purposes, but also established the basis for the development of

molecular microbial ecology (Amann et al., 1995). Due to that reason, in parallel to the sequencing of new cultured isolates, ecologists generated important numbers of environmental sequences originated from uncultured environmental microorganisms. The increase was significatively larger than that of cultured organisms, and surpassed them in orders of magnitude (Figure 1), being nowadays almost 3,000,000 sequences, 99% of them just environmental (www.arb-silva.de).

What is a species (concept):

Microbiologists regard species as a "monophyletic group of isolates showing high genomic coherence and that share a large set of independent phenotypic characters"

How do microbiologists circumscribe a species (definition):

In order to properly describe a species, one has to make the effort in isolation as many strains as possible of the same group. When it is not possible, then the species description will be a single organism description. To embrace a group of isolates as a single species one has to: (i) show that all organisms are monophyletic (using the analysis of gene sequences as 16S rRNA, or alternative markers with phylogenetic signal); (ii) show that their genomes are closely related, in general >96% ANI for intraspecies circumscriptions and <94% for interspecies discrimination; and (iii) show with as many independent phenotypic (metabolic, enzymatic, chemotaxonomic…) traits that the group of organisms share a phenotypic property that makes them different from the closest relatives.

Microbial taxonomy benefited importantly from all such developments, and have modified the way to circumscribe species and higher taxa based on that molecular approaches. In this regard, the current concept of species for prokaryotes considers this unit to be a "monophyletic group of isolates that show high genomic coherence, and that share a large set of independent phenotypic characters" (Rosselló-Mora and Amann, 2001). The main disadvantage of the taxonomic shema for prokaryotes is that for classification purposes it is necessary to isolate the organisms as pure culture in the laboratory. It is not yet possible to construct a reliable and predictable taxonomy based on not yet cultivated microorganisms. However, another of the important achievements in taxonomy is the obligatory deposit of at least one designated strain (the so called type strain) in two international culture collections of different countries. This deposit guarantees the perpetuation of living material as well as the availability for any scientist.

In this regard, there are two important developments towards the improvement of the taxonomic activities of classification. In first instance the generation of a 16S rRNA database just containing type strain sequences, the so called LTP (www.arb-silva.de/projects/living-tree). This database was generated in order to collect from the enormous 16S rRNA repositories the best sequences representing the validly named species (Yarza et al., 2008). This database, that is several times updated per year, has been manually curated in order to correct the entry mistakes that often appear in the submissions. In more than 12% of the cases, the specific name in the information fields of the sequence entries are wrong, often due to the lack of updates done by the respectie authors. The LTP arranges the sequences as well using the universal alignment established in the ARB software package (Ludwig et al., 2004), that takes into account the secondary structure of the ribosomal RNA, and helps in the recognition of homologous positions.

A second important development for taxonomy has been the search for an alternative to the DDH experiments. This technique is for some researchers cumbersome, but the most important pitfall is actually that the data generated cannot be stored in cumulative databases as the gene sequences. For this, Konstantinidis and Tiedje (2005) conceived the parameter ANI (Average Nucleotide Identity),

which is the mean identity value among all homologous stretches shared by two genomes. Our group (Richter and Rosselló-Móra, 2009) generated the program package JSpecies (www.imedea.uib.es/jspecies) to calculate this value as well as the regression of the tetranucleotide signatures among two different genomes. The results of the application of ANI nicely showed that the the 70% threshold established for DDH comparisons matched ANI values ranging between 94 to 96%. In addition, we could show that there was a very nice correlation between ANI and the regression values of the tetranucleotide signatures, for which when pairwise genomes show regression above 0.999, then they can be considered of the same species.

ANI is predicted to substitute DDH in the very near future. Most probably, its use will influence as well the minimal standards for species circumscriptions. Any type strain of any new species will be fully sequenced in the future, and the complete dataset of genomes of validly published species names would be an excellent target to the recognition of any new species. The advantage of ANI is that it does not require full genome sequences for comparisons. Two genomes randomly sequenced would render plausible ANI results with just 20% of each sequence. Moreover, when the complete dataset of type strain genomes is achieved, just the random sequencing of about 4% of a genome would be enough to obtain an ANI value that indicates whether represents or not a new species. It is to foreseen that genome sequencing will become a routine in the future, and probably compulsory for any new description as it is now the 16S rRNA.

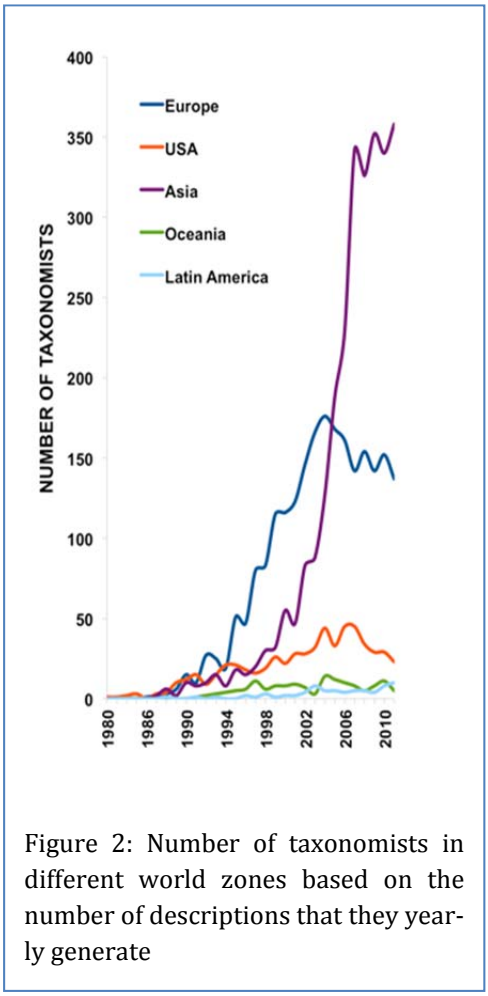

Figure 2: Number of taxonomists in different world zones based on the number of descriptions that they yearly generate

Here we have seen how the technological developments (especially the sequencing of the 16S rRNA) have improved the rate of descriptions. However, during the last decade there has been a shift on the habits and nationalities of taxonomists describing species (Tamames and Rosselló-Móra, 2012). Europe and USA have been leading microbial taxonomy during most of its existence. Actually, scientists from these nations have developed the current classification scheme as well as the different criteria for classification. However, since the early XXIst century, the asian countries (China, Japan and South Korea) have taken the leadership in classifying new organisms, whereas in Europe and USA, the number of active taxonomists seem to be in decline (Figure 2). The loss of expertise in Western countries might not only cause a loss of knowledge on how to interpret microbial diversity, but also the reduction in the deposit of isolates in their own culture collections. Altogether, this decline can be dangerous for both, the basic research knowledge and the potential to access classified organisms in strain collections. The reasons for this might be found in the low access to research funds.

Current challenges in microbial taxonomy: One of the major challenges is the access to proper funds to support the know how in those nations where taxonomists seem to be in decline. In any case, another challenge is the achievement that for any new species description, the full genome sequence deposit is a compulsory requirement. Such sequences do not only have taxonomic value, but also have the potential of predicting metabolic and genetic traits of the isolates.

Future needs: Microbial taxonomy will increasingly be dependent on the development of high throughput methods to reveal genetic and metabolic diversity (all so called –omics). The expansion of such techniques will require interactive databases able to match their results and automatize the identification of new isolates by means of genetic or phenotypic data.

The relevance of taxonomy for the society: Taxonomy deals with the understanding of the biological diversity present in the biosphere as well as with the generation of a classification system that is of universal use, operational and predictive. A universally useful system means that it should be usable by any discipline in microbiology, and serves as a communication tool among disciplines. An operational system means that non-experts should be able to understand and use it. And a predictive system means that the identification of an organism as a member of an existing taxon will simultaneously indicate its genetic, phenotypic and ecological properties with low degree of uncertainty. Society benefits from taxonomy by means of the knowledge of the extent of diversity of the microorganisms that in some cases can promote the development of biotechnological tools with relevance in the pharmaceutical or food industries, among others.

Literature

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol Rev 59: 143-169

Konstantinidis, K., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 102**:** 2567–2592.

Ludwig,W., Strunk,O., Westram,R., Richter,L., Meier,H., Yadhukumar, Buchner,A., Lai,T., Steppi,S., Jobb,G., Forster,W., Brettske,I., Gerber,S., Ginhart,A.W., Gross,O., Grumann,S., Hermann,S., Jost,R., Konig,A., Liss,T., Lussmann,R., May,M., Nonhoff,B., Reichel,B., Strehlow,R., Stamatakis,A. ,Stuckmann,N., Vilbig,A., Lenke,M., Ludwig,T., Bode,A., Schleifer,K.-H.(2004) ARB: a software environment for sequence data. Nucleic Acids Res .32:1363-1371.

Rosselló-Móra, R. (2012) Towards a taxonomy of *Bacteria* and *Archaea* based on interactive and cumulative data repositories. Environ. Microbiol. 14:318-334

Rosselló-Móra, R., and Amann, R.I. (2001) The species concept for prokaryotes. FEMS Microbiol Rev25**:** 39–67

Tamames, J., and Rosselló-Móra, R. (2012) On the fitness of microbial taxonomy. Trends in Microbiol. IN PRESS

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, L., *et al*. (1987) Report of the *adhoc* committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol 37**:** 463–464

Woese,C.R. (1987) Bacterial evolution. Microbiol. Rev. 51,2,221-271

Yarza,P., Richter,M., Peplies,J., Euzéby,J., Amann,R., Schleifer,K.-H., Ludwig,W., Glöckner,F.O., Rosselló-Móra,R. (2008) The All-Species Living Tree Project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst .Appl. Microbiol. 31:241–250.

The intestinal microbiome and human health: a new challenge for taxonomy of metagenomic data

– Bärbel Stecher

Max von Pettenkofer-Institute for Hygiene and Clinical Microbiology of Ludwig Maximilians-University Munich, Department Bacteriology, Munich, Germany

Introduction

The mammalian gastrointestinal tract contains an immense number and diversity of microorganisms, and assembly, composition and interspecies interactions within this unique microbial ecosystem are still insufficiently explored. Resistance of the host to bacterial infections is critically dependent on microbiota integrity (Sekirov *et al.* 2010; Stecher & Hardt 2010) and infections caused by several major human pathogens (*Salmonella* spp., EHEC, VRE and *Clostridium difficile*) are efficiently prevented by the host-intrinsic microbiota. Furthermore, pathogens can interact with members of the commensal microbiota and acquire virulence factors and antibiotic resistances by horizontal gene transfer. The majority of the human-associated commensal bacteria have not been cultured yet and taxonomical characterization is mostly incomplete at the current stage. Therefore, it is important to get further insights into the composition, species content, functionality and stability of the human microbiota.

So far, comprehensive surveys including hundreds of human individuals have investigated microbial diversity at different human body habitats as well as their functional and phylogenetic stability over time (Arumugam *et al.* 2011; The Human Microbiome Project Consortium 2012; Yatsunenko *et al.* 2012). The rapidly emerging field of microbiome studies offers investigators a large choice of sequencing methods as well as bioinformatics tools for determining the content of microorganisms in a sample. Unfortunately however, metagenome analysis and 16S rRNA fingerprinting are insufficient to measure actual species dynamics and variability (i.e. by horizontal gene transfer) in the gut.

Highly parallel next generation sequencing technologies have allowed us gaining deeper insights into the genetic makeup of human-associated microbiota and characterizing the diversity and stability of this special microbial community over time and in between different human individuals (Weinstock 2012). Apparently, inter- and intra-individual diversity on a phylogenetic level can be rather high whereas, surprisingly, the key functions (i.e. the presence of genes encoding specific pathways) are conserved. Recent studies investigated how diet, life-style, antibiotics and diseases can impact on microbiome composition and function (Dethlefsen *et al.* 2008; Yatsunenko *et al.* 2012). Importantly, immune deficiency, antibiotic use, diet and infections can provoke severe microbiota perturbations leading to state of “dysbiosis” (i.e. a pathologically altered microbiome) with severe consequences for human health (Hooper, Littman & Macpherson 2012). Knowing the basic parameters on structure and stability of the human microbiome, in the next step, one can continue addressing how microbiota alterations correlate with the incidence and development of metabolic diseases, carcinogenesis, cardiovascular and infectious diseases.

What are the current main challenges in your research area?

Hitherto, the majority of studies analyzed microbial community composition by 16S rRNA amplicon sequencing (Huse *et al.* 2008). This is a valid approach which can be performed in high-throughput at relative low cost but involves certain drawbacks as compared to metagenome analysis: PCR amplification introduces a strong bias to the bacterial community represented within the amplified sequences which is i.e. due to primer specificity and multi-template PCR amplification artifacts (Kanagawa 2003). Moreover, phylogenetic information within the 16S rRNA gene is mostly insufficient to readily differentiate bacteria on a species- (e.g. *E. coli* and *Shigella* spp.) and strain-level (i.e. *E. coli* pathotypes). In current studies, the short lengths of the amplicons merely enables bacterial discrimination on family or genus level.

Metagenome sequencing captures bulk microbial, viral and host DNA. However, the assembly of individual bacterial genomes from such data is still quite challenging due to the short sequence reads which is insufficient for genome assembly (Weinstock 2012). This implies that the genetic content (i.e. genome) of a single "transmissible" unit (one bacterial strain) cannot be readily predicted from the current metagenomic data.

Full genome assembly of individual strains from metagenomes would be highly desirable for the following reasons:

(1) Assembly of individual bacterial genomes would permit the analysis of the temporal evolution of individual bacterial strains and get insights into the strains' potential to accumulate mutations as well as to exchange genetic material via horizontal gene transfer (HGT), i.e. in between pathogenic and harmless gut inhabitants (Zaneveld, Nemergut & Knight 2008).

(2) Certain genes (i.e. encoding antibiotic resistance cassettes, toxins) only imply an increased health risk if they are encoded in a specific strain background (i.e. a pathogen or opportunist). This cannot readily be integrated from metagenomic data. In case of plasmid-encoded resistances, it cannot be determined at all.

(3) It is important to address if the carriage of virulence factors and antibiotic resistances by specific commensal strains among the intrinsic microbiota correlates with an elevated risk of spreading the genes to pathogens by HGT ("commensal vectors") and thus speed up the evolution of pathogens with increased virulence.

(4) This type of data can help revealing how often particular strains are transmitted in between different individuals (i.e. partners, family, and persons within the same household or work place)?

(5) If "carriage" of a certain microbial signature was associated with an increased risk of developing certain types of diseases it would be of major importance to assess the infection risk coming from those individuals. To prevent transmission, the subset of potentially harmful strains has to be identified and thoroughly characterized to assess their contagiousness.

(6) Near complete assembly of bacterial genomes can aid validating the success of therapeutic measures targeting the human microbiome (i.e. pathogen eradication, pre- and probiotic therapy) using metagenomic analysis.

What needs do you see in the next future for your research area?

Generation of reference genome sequences: Until recently, sequencing of bacterial genomes had been rather costly and laborious. After the advent of next-generation sequencing technology the price to pay for a bacterial shotgun genome sequence dropped below 10,000 € for sequencing costs and initiatives have been launched to provide **reference-genome-sequences** of cultured human and animal bacterial isolates (Human Microbiome project; GEBA project). This resource of fully assembled genomes will be extremely valuable for studying genome structure and evolution of commensal bacteria. The database will also allow comparative analysis of strains derived from different host organisms (e.g. humans, mammals, birds) and help identifying the mechanisms driving host specificity, functionality and adaptation of commensal bacteria (Chung *et al.* 2012).

Further development of metagenome analysis tools: Development of bioinformatics software and analysis strategies as well as computer capacity is constantly lagging behind the exponential increase of sequence data being daily generated by sequencing platforms. Thus, further development of bioinformatics software tools for sequence data handling, metagenome analysis, genome assembly as well as user-friendly analysis pipelines should be a central focus of future research initiatives in this area.

Development and improvement of preclinical animal model systems. A number of studies have identified the intestinal microbiota as central player of human health and disease. Thus, therapeutic manipulation of the microbiota is a major focus of clinical research. For development and assessment of new therapies (i.e. probiotics, antibiotics, fecal transplantation), simplified preclinical model systems (e.g. gnotobiotic mouse models) are urgently needed. Gnotobiotic mouse models, in combination with genetically engineered mice are instrumental in addressing the role of bacteria or their products in health and disease (Smith, McCoy & Macpherson 2007). Mechanistic links have been revealed between targeted microbiota manipulation and diseases including obesity, multiple sclerosis, inflammatory bowel disease and others (Balish & Warner 2002; Turnbaugh *et al.* 2006; Lee *et al.* 2010). Gnotobiotic “humanized” models have been established, i.e. rodents colonized with a defined human-derived microbiota (Faith *et al.* 2010). Those model systems will also be of major importance for basic research to gain more insights into the speed of evolution of the human microbiota and how this may be influenced by environmental parameters (e.g. diet, travelling, antibiotics and diseases). This is of major importance as to date, the actual frequency of genetic exchange, its hotspots and limitations in terms of species boundaries and contributions to ecosystem functionality and spread of antibiotic resistances and virulence factors are largely unclear.

Why taxonomy is relevant to society?

Maintaining a mutually beneficial relationship with our commensal microbial community is of central importance for health. Exploiting the intestinal microbiota in clinical applications as therapeutic target as well as therapeutic agent bears great potential. Human fecal transplants are already being used to cure infectious disease like recurrent *Clostridium difficile* infections (Palmer 2011). Yet, transplantation of largely uncharacterized, freshly harvested fecal slurries implies a certain risk of transmission of infections and other unwanted effects. To this end, development of novel, well-characterized probiotics (i.e. based on human microbial isolates) is highly desirable in future therapeutic applications. Since the vast number of species present in the human gut has never been cul-

tured before it is unclear, how divergent the single species /strains are and how fast they evolve. Taxonomic characterization and classification is a highly relevant issue i.e. for patenting those strains as well as for conducting safety assessment of novel probiotics in human clinical trials (i.e. specific detection of the presence of the strains).

References

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Dore, J., Consortium, M., Weissenbach, J., Ehrlich, S.D., Bork, P., Antolin, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y. & Zeller, G. (2011) Enterotypes of the human gut microbiome. *Nature*.

Balish, E. & Warner, T. (2002) Enterococcus faecalis induces inflammatory bowel disease in interleukin-10 knockout mice. *Am J Pathol,* **160,** 2253-2257.

Chung, H., Pamp, S.J., Hill, J.A., Surana, N.K., Edelman, S.M., Troy, E.B., Reading, N.C., Villablanca, E.J., Wang, S., Mora, J.R., Umesaki, Y., Mathis, D., Benoist, C., Relman, D.A. & Kasper, D.L. (2012) Gut immune maturation depends on colonization with a host-specific microbiota. *Cell,* **149,** 1578-1593.

Dethlefsen, L., Huse, S., Sogin, M.L. & Relman, D.A. (2008) The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biol,* **6,** e280.

Faith, J.J., Rey, F.E., O'Donnell, D., Karlsson, M., McNulty, N.P., Kallstrom, G., Goodman, A.L. & Gordon, J.I. (2010) Creating and characterizing communities of human gut microbes in gnotobiotic mice. *ISME J,* **4,** 1094-1098.

Hooper, L.V., Littman, D.R. & Macpherson, A.J. (2012) Interactions between the microbiota and the immune system. *Science,* **336,** 1268-1273.

Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A. & Sogin, M.L. (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet,* **4,** e1000255.

Kanagawa, T. (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng,* **96,** 317-323.

Lee, Y.K., Menezes, J.S., Umesaki, Y. & Mazmanian, S.K. (2010) Microbes and Health Sackler Colloquium: Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc Natl Acad Sci U S A*.

Palmer, R. (2011) Fecal matters. *Nat Med,* **17,** 150-152.

Sekirov, I., Russell, S.L., Antunes, L.C. & Finlay, B.B. (2010) Gut microbiota in health and disease. *Physiol Rev,* **90,** 859-904.

Smith, K., McCoy, K.D. & Macpherson, A.J. (2007) Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin Immunol,* **19,** 59-69.

Stecher, B. & Hardt, W.D. (2010) Mechanisms controlling pathogen colonization of the gut. *Curr Opin Microbiol*.

The Human Microbiome Project Consortium (2012) A framework for human microbiome research. *Nature,* **486,** 215-221.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. & Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature,* **444,** 1027-1031.

Weinstock, G.M. (2012) Genomic approaches to studying the human microbiota. *Nature,* **489,** 250-256.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R. & Gordon, J.I. (2012) Human gut microbiome viewed across age and geography. *Nature,* **486,** 222-227.

Zaneveld, J.R., Nemergut, D.R. & Knight, R. (2008) Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology,* **154,** 1-15.

Characterizing species boundaries and species histories in closely related fungi using comparative population genomic approaches

– Eva H. Stukenbrock

Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

Introduction

The correct identification and characterization of fungal species is essential in many disciplines of applied and basic research. Fungi are widely explored in biotechnology where one goal is the discovery of new species for the production of beneficial metabolites. In medical research and plant pathology an essential task is the correct identification and characterization of disease-causing organisms with the aim of developing appropriate treatments and management strategies. In biological sciences, species recognition has become an essential component in biodiversity research and conservation biology for the quantitative and qualitative description of species richness. I here propose an approach based on genomic and evolutionary biology to support the identification of evolutionary independent lineages (species). Furthermore, I suggest that the recognition of within-species variation should be considered in taxonomic species descriptions. The definition of a species used here refers to a single evolutionary lineage of organisms that maintains its integrity with respect to other lineages through both time and space.

Species recognition in fungi

Morphological and biological species recognition in fungi is challenged by the lack of comparable morphological structures and the absence of reproductive barriers between many closely related species. Phylogenetic species recognition using information from independent sequence loci may also fail to correctly identify evolutionary independent lineages due to the lack of sufficient time for mutations to become fixed, or lineage sorting to occur. To improve the distinction of closely related species Taylor et al. proposed the use of a Genealogical Concordance Phylogenetic Species Recognition (GCPSR) concept based on concordance between gene genealogies of independent loci taylor (Taylor *et al.* 2000). The GCPSR term builds on population genetics concepts where conflicts between gene trees arise from recombination between individuals. The transition from concordance to conflict thereby determines the limit of species that are expected to be reproductively isolated. A drawback of this approach is that the transition from concordance to conflict does not occur for clonal species. Since many fungal species are known to propagate almost exclusively by asexual reproduction conflicting gene genealogies cannot be used to delimit species boundaries. Another problem with the GCPSR approach is that conflicts between gene genealogies of closely related species may be due to incomplete sorting of polymorphisms in the ancestral species and not due to recombination between present day species.

Genomics: New tools in taxonomy

Advances in genome sequencing and bioinformatics approaches have opened a new avenue to assess lineage divergence and species boundaries in fungi. Comparative genomics of two or more spe-

cies allow (i) the identification of species-specific traits, and (ii) comparisons of genome contents and structure. This is particularly relevant for the distinction of closely related clonal species, where there can be a continuum of genetic differentiation and the recognition of genetic isolation is more difficult to assess. Species-specific traits may be genomic regions that are unique to a particular species or strongly modified compared to the homologous loci in close relatives. These can include horizontally transferred sequences (Friesen *et al.* 2006), isochores (Rouxel *et al.* 2011), or duplicated genomic regions (Scannell, Butler & Wolfe 2007). The identification of such species-specific traits may support the delimitation of species boundaries and/or add information to their biological interpretation, if these traits confer morphological phenotypic differences or even confer reproductive isolation between lineages.

Comparative genomics also allows us to infer parameters related to species histories. Whole genome coalescence approaches are new tools to estimate genomic relationships and speciation times (Hobolth *et al.* 2007), as well as ancestral recombination maps and incomplete lineages sorting (Dutheil *et al.* 2009) (Figure 1). Recently, a method has been developed for inferences of genome wide ancestral migration rates allowing fine-scale insight into the emergence of reproductive boundaries and past introgression during the divergence of species (Mailund et al, in revision). These coalescence approaches provide insight into species evolution but also serve as tools to characterize species delimitations (Fujita *et al.* 2012). Speciation times can be computed with posterior probabilities providing robust and standardized measures of the evolutionary independence of species.

Another genome-view of speciation is through the inferences of incomplete lineages sorting in genome alignments of closely related species (Figure 1) (Dutheil *et al.* 2009). With incomplete lineage sorting the genealogy of sites differs across the genome. This can be due to stochastic processes, natural selection or recombination events in the common ancestor of present day species. In phylogenetics, this results in incongruence between phylogenetic trees reconstructed from distinct genomic regions. Genome-wide coalescence approaches allow the genealogy of each site in the genome alignment of closely related species to be computed (Dutheil *et al.* 2009). The resulting pattern from a genome wide analysis will be a mosaic of genealogies, some of which are congruent with the canonical species tree and others, which will have an alternative topology. The proportion of sites with incomplete lineage sorting can be computed and can serve as a different quantitative measure in the recognition of evolutionary species and as an indicator for the rate of species events.

While between species comparative genomics allows the identification of species specific traits, within species variability should still be accounted for. It is known that individuals of a species may show a range of different phenotypes. Still at the phylogenetic level they can be characterized as one species. Population genomic sequencing will support the recognition and quantification of within species variability (Stukenbrock *et al.* 2011; Stukenbrock *et al.* 2012). It should be noted that the level of within species variation might vary significantly depending on the species history or life history traits of the species.

What are the challenges in your research area?

In terms of species recognition and species characterization, a challenge is to account for within species variation. We observe in some cases that the within species differentiation exceeds the described differentiation between species for certain characters. While within-species variability may

blur the limit of species boundaries, coalescence based approaches and phylogenetic approaches using the GCPSR concept can support the recognition of evolutionary independent lineages. The availability of genome and transcriptome data will in the future furthermore allow us to identify genetic traits responsible for phenotypic and morphological characters and possibly the genetic basis of species boundaries. A goal should be the integration of information from genomic and phenotypic traits from multiple individuals to link genetic variability to the observed phenotypic variability.

What needs do you see in the next future for your research area?

In the near future the access to genome data and the sequencing of multiple individuals will no longer be a limiting factor. Genome, transcriptome and metagenome databases will be valuable resources for taxonomic research. However, the analyses of this data and the development and testing of appropriate analytical tools is still a challenging step. Coalescence based methods have so far only in a few studies been implemented as a tool for species delimitation (see Fujita *et al.* 2012 for review), and the potential and further integration of such approaches in taxonomy needs to be discussed.

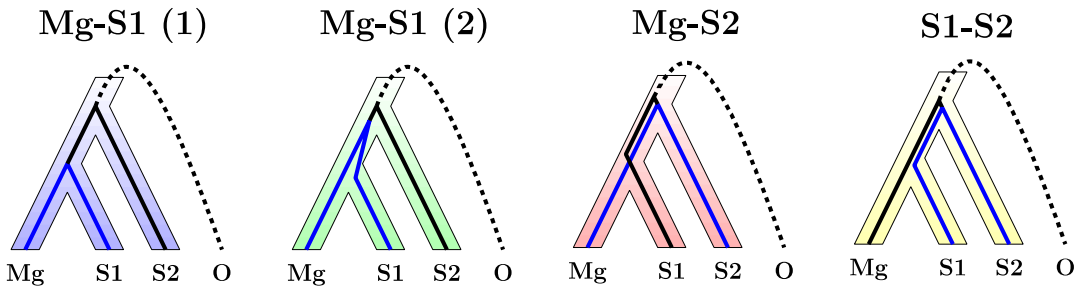

Figure 1: A graphical example of incomplete lineages sorting. With incomplete lineage sorting the genealogy differs between independent loci or along the genome. With the first genealogy Mg-S1 the tree corresponds to the canonical species tree where S1 is the closest relative of Mg. The second genealogy still corresponds to the species tree however, the divergence of the Mg and S1 alleles occurred already in the common ancestor of Mg-S1-S2. For some sites in the genome the genealogy however does not correspond to the species tree. With the alternative scenarios the most closely related sequences can be Mg and S2 or S1 and S2. Coalescence based methods allow us to infer the genome wide proportion of sites with alternative genealogies (incomplete lineages sorting) (Dutheil *et al.* 2009). Figure from (Stukenbrock *et al.* 2011).

Why taxonomy is relevant to society

Taxonomy of fungal research is highly relevant to society through the importance of fungi in biotechnology, medicine, agronomy and ecosystem conservation.

In the aim of developing new or improved products, biotechnology companies search for hitherto unknown fungal species with the capacity of producing components with desired characteristics. Thus fungi are sampled from habitats such as rainforests and artic tundra. These fungi must obviously be taxonomically classified.

In plant pathology as well as medicine the correct identification of fungal pathogens is essential for appropriate disease treatment. In plant pathology an important challenge is the global dissemination of pathogens for example by transport of infected plant material. Taxonomic recognition of such pathogens is important to track migration routes and foresee further epidemic spreads.

Also beneficial fungi such a mychorrhizal species are currently being explored as bio fertilizers and the correct identification of these and native mycorrhiza species is critical to evaluate the effect on plant growth and the impact of introduced species on the native microbial community.

References

Dutheil, J.Y., Ganapathy, G., Hobolth, A., Mailund, T., Uyenoyama, M.K. & Schierup, M.H. (2009) Ancestral population genomics: The coalescent hidden Markov model approach. *Genetics,* **183,** 259-274.

Friesen, T.L., Stukenbrock, E.H., Liu, Z., Meinhardt, S., Ling, H., Faris, J.D., Rasmussen, J.B., Solomon, P.S., McDonald, B.A. & Oliver, R.P. (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics,* **38,** 953-956.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A. & Moritz, C. (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution,* **27,** 480-488.

Hobolth, A., Christensen, O.F., Mailund, T. & Schierup, M.H. (2007) Genomic relationships and speciation times of Human, Chimpanzee, and Gorilla inferred from a coalescent hidden Markov model. *PLoS Genetics,* **3,** e7.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A.J., Ciuffetti, L.M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S.B., Gout, L., Glaser, N., Linglin, J., Kema, G.H.J., Lapalu, N., Lawrence, C.B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C.L., Simon, A., Spatafora, J.W., Stachowiak, A., Turgeon, B.G., Tyler, B.M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R.P., Wincker, P., Balesdent, M.-H. & Howlett, B.J. (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat Commun,* **2,** 202.

Scannell, D.R., Butler, G. & Wolfe, K.H. (2007) Yeast genome evolution—the origin of the species. *Yeast,* **24,** 929-942.

Stukenbrock, E.H., Bataillon, T., Dutheil, J.Y., Hansen, T.T., Li, R., Zala, M., McDonald, B.A., Wang, J. & Schierup, M.H. (2011) The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research,* **21,** 2157-2166.

Stukenbrock, E.H., Christiansen, F.B., Hansen, T.T., Dutheil, J.Y. & Schierup, M.H. (2012) Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proc Natl Acad Sci,* **109** (27), 10954-10959 .

Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S. & Fisher, M.C. (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology,* **31,** 21-32.

(Gen-)Omics and bacterial pathogens: Impact and Potential

– Sebastian Suerbaum

Hannover Medical School, Institute for Medical Microbiology and Hospital Epidemiology, Hannover, Germany

Introduction

The interest in bacteria causing human infectious diseases has been a major driving force of genomic research in bacteriology. Most of the first published complete bacterial genome sequences were from human pathogens, such as *Haemophilus influenzae* (Fleischmann et al. 1995), *Mycoplasma* sp. (Fraser *et al.* 1995), or *Helicobacter pylori* (Tomb *et al.* 1997; Alm *et al.* 1999), and interest in pathogens has likewise fuelled other genome-based technologies, such as transcriptomics and proteomics. The human chronic gastric pathogen, *Helicobacter pylori*, whose genomic evolution is a major focus of my laboratory, has become a paradigm for the potential of genomic analyses to improve our understanding of the evolution of a major pathogen together with its human host as well as the adaptation of the bacterial population to an individual carrier during chronic infection (Falush *et al.* 2003; Linz *et al.* 2007; Suerbaum & Josenhans 2007; Kennemann *et al.* 2011). Comparative genome analyses of dozens to hundreds of bacterial strains have recently become one of the most powerful tools to decipher the evolution and spread of pathogens, such as the *Yersinia pestis*, the cause of plague (Morelli *et al.* 2010), *Mycobacterium tuberculosis* roetzer (Roetzer *et al.* 2013), or the major nosocomial pathogens, *Staphylococcus aureus* (Holden *et al.* 2013) and *Clostridium difficile* (Didelot *et al.* 2012).

The impact of genomics on the practice of diagnostic and clinical microbiology has so far been limited. By contrast, the application of proteome-based technologies has recently revolutionized the identification of pathogenic bacteria in high-throughput diagnostic laboratories (Bizzini & Greub 2010). Bacterial genome sequencing has recently achieved wide attention in conjunction with outbreaks such as the cholera outbreak in Haiti (Chin *et al.* 2011), and the *Escherichia coli* O104:H4 outbreak in Germany (Brzuszkiewicz *et al.* 2011; Mellmann *et al.* 2011; Rohde *et al.* 2011). The recent advances in our understanding of the role of the human microbiota (in particular, the gut microbiota) in health and disease have spurred a flurry of studies with the goal to identify microbiota-based diagnostic and predictive markers for human disease and raised hopes that modification of the microbiota can improve resistance to pathogen intrusion and metabolic diseases (Turnbaugh *et al.* 2009; Lozupone *et al.* 2012; Maurice & Turnbaugh 2013). Finally, genome-based approaches to vaccine development ('reverse vaccinology') have yielded promising results and the first vaccine designed by this approach, directed against serotype B meningococci, has now been licensed in several countries (Sette & Rappuoli 2010).

Theses

- Interest in bacteria with relevance to human health has been a major driving force of the (gen-)omic revolution.

- Understanding the taxonomy of pathogenic bacteria and related non-pathogenic species is a prerequisite for accurate diagnostics and reliable interpretation of susceptibility data.
- Changes in the taxonomic classification of medically relevant bacteria ("name changes") carry risks for patient care that need to be anticipated.
- The application of high throughput genome sequencing to large and representative strain collections for all medically relevant bacteria has enormous potential to increase our understanding of pathogen evolution, transmission, and pathogenesis. In addition, genomics is the basis of 'reverse vaccinology', a rational approach to the development of vaccines against pathogens where conventional approaches to vaccine development have failed until now.
- Research about the role of the microbiota in human health and disease has enormous potential but the analysis of the required large numbers of samples from human individuals and experimental model systems still poses a major bottleneck. Input from taxonomists and bioinformaticians is quintessential to the success of such projects.

References

Alm, R.A., Ling, L.-S.L., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F. & Trust, T.J. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature,* **397,** 176-180.

Bizzini, A. & Greub, G. (2010) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. *Clinical Microbiology and Infection,* **16,** 1614-1619.

Brzuszkiewicz, E., Thürmer, A., Schuldes, J., Leimbach, A., Liesegang, H., Meyer, F.-D., Boelter, J., Petersen, H., Gottschalk, G. & Daniel, R. (2011) Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Archives of Microbiology,* **193,** 883-891.

Chin, C.-S., Sorenson, J., Harris, J.B., Robins, W.P., Charles, R.C., Jean-Charles, R.R., Bullard, J., Webster, D.R., Kasarskis, A., Peluso, P., Paxinos, E.E., Yamaichi, Y., Calderwood, S.B., Mekalanos, J.J., Schadt, E.E. & Waldor, M.K. (2011) The origin of the Haitian cholera outbreak strain. *New England Journal of Medicine,* **364,** 33-42.

Didelot, X., Eyre, D., Cule, M., Ip, C., Ansari, M., Griffiths, D., Vaughan, A., O'Connor, L., Golubchik, T., Batty, E., Piazza, P., Wilson, D., Bowden, R., Donnelly, P., Dingle, K., Wilcox, M., Walker, A., Crook, D., A Peto, T. & Harding, R. (2012) Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biology,* **13,** R118.

Falush, D., Wirth, T., Linz, B., Pritchard, J.K., Stephens, M., Kidd, M., Blaser, M.J., Graham, D.Y., Vacher, S., Perez-Perez, G.I., Yamaoka, Y., Mégraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M. & Suerbaum, S. (2003) Traces of human migrations in *Helicobacter pylori* populations. *Science,* **299,** 1582-1585.

Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., Bult, C., Tomb, J., Dougherty, B., Merrick, J. & al., e. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* **269,** 496-512.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A. & Venter, J.C. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science,* **270,** 397-404.

Holden, M.T.G., Hsu, L.-Y., Kurt, K., Weinert, L.A., Mather, A.E., Harris, S.R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Žemličková, H., Coombs, G., Kearns, A.M., Hill, R.L.R., Edgeworth, J., Gould, I., Gant, V., Cooke, J., Edwards, G.F., McAdam, P.R., Templeton, K.E., McCann, A., Zhou, Z., Castillo-Ramírez, S., Feil, E.J., Hudson, L.O., Enright, M.C., Balloux, F., Aanensen, D.M., Spratt, B.G., Fitzgerald, J.R., Parkhill, J., Achtman, M., Bentley, S.D. & Nübel, U. (2013) A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Research,* **23,** 653-664.

Kennemann, L., Didelot, X., Aebischer, T., Kuhn, S., Drescher, B., Droege, M., Reinhardt, R., Correa, P., Meyer, T.F., Josenhans, C., Falush, D. & Suerbaum, S. (2011) *Helicobacter pylori* genome evolution during human infection. *Proceedings of the National Academy of Sciences*.

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., Yamaoka, Y., Graham, D.Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S. & Achtman, M. (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature,* **445,** 915-918.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. & Knight, R. (2012) Diversity, stability and resilience of the human gut microbiota. *Nature,* **489,** 220-230.

Maurice, C.F. & Turnbaugh, P.J. (2013) Quantifying the metabolic activities of human-associated microbial communities across multiple ecological scales. *FEMS Microbiology Reviews,* **37,** 830-848.

Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S.F., Henkhaus, J.K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P.M., Moore, R.L., Guenther, S., Rothberg, J.M. & Karch, H. (2011) Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE,* **6,** e22751.

Morelli, G., Song, Y., Mazzoni, C.J., Eppinger, M., Roumagnac, P., Wagner, D.M., Feldkamp, M., Kusecek, B., Vogler, A.J., Li, Y., Cui, Y., Thomson, N.R., Jombart, T., Leblois, R., Lichtner, P., Rahalison, L., Petersen, J.M., Balloux, F., Keim, P., Wirth, T., Ravel, J., Yang, R., Carniel, E. & Achtman, M. (2010) *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nature Genetics,* **42,** 1140-1143.

Roetzer, A., Diel, R., Kohl, T.A., Rückert, C., Nübel, U., Blom, J., Wirth, T., Jaenicke, S., Schuback, S., Rüsch-Gerdes, S., Supply, P., Kalinowski, J. & Niemann, S. (2013) Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: A longitudinal molecular epidemiological study. *PLoS Medicine,* **10,** e1001387.

Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N.J., Hentschke, M., Chen, W., Pu, F., Peng, Y., Li, J., Xi, F., Li, S., Li, Y., Zhang, Z., Yang, X., Zhao, M., Wang, P., Guan, Y., Cen, Z., Zhao, X., Christner, M., Kobbe, R., Loos, S., Oh, J., Yang, L., Danchin, A., Gao, G.F., Song, Y., Li, Y., Yang, H., Wang, J., Xu, J., Pallen, M.J., Wang, J., Aepfelbacher, M. & Yang, R. (2011) Open-Source genomic analysis of shiga-toxin–producing *E. coli* O104:H4. *New England Journal of Medicine,* **365,** 718-724.

Sette, A. & Rappuoli, R. (2010) Reverse vaccinology: Developing vaccines in the era of genomics. *Immunity,* **33,** 530-541.

Suerbaum, S. & Josenhans, C. (2007) *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nature Reviews Microbiology,* **5,** 441-452.

Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzegerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. & Venter, J.C. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature,* **388,** 539-547.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R. & Gordon, J.I. (2009) A core gut microbiome in obese and lean twins. *Nature,* **457,** 480-484.

Textbeiträge Zoologie-Workshop

Wolbachia and molecular barcoding - conceptions and misconceptions

– Christoph Bleidorn

University of Leipzig, Institute for Biology, Molecular Evolution and Systematics of Animals, Leipzig, Germany

The aim of the Barcode of Life initiative is to provide a standardized genetic marker system for species identification. In the case of animals, a 648 bp region of the mitochondrial cytochrome oxidase subunit I gene (cox1) is used as a barcode (Hebert et al., 2003). However, the reliance on a single, maternally inherited marker provoked a controversy regarding the limits of such an approach.

Insects are thought to represent the most species rich animal group. Consequently, several barcoding initiatives aim to build large resources for selected insect taxa. E.g., several studies investigated Lepidoptera as a flagship taxon to estimate the reliability of the barcoding approach. The success of species identification is usually estimated by comparing the results with current taxonomy and mismatches are declared as cases of so far unrecognized cryptic species or "bad taxonomy". However, it would be straight forward to use nuclear genetic marker as a reference, as mitochondrial and nuclear gene trees might yield incongruent results. Reasons include the reduced effective population size of strictly maternal inherited genes, introgression, inconsistent mutation rate, pseudogenization, or heteroplasmy (Galtier et al. 2009). Furthermore, in the case of terrestrial arthropods the presence of cytoplasmatic bacteria can dramatically alter inheritance patterns of mitochondrial genes. Different lineages of reproductive parasites are capable of inducing such effects, e.g., *Cardinium* or *Rickettsia.* The by far best known and most widespread examples are Alpha-proteobacteria of the genus *Wolbachia*. Four mechanisms of host manipulations are distinguished: (1) feminization of genotypic males, (2) induction of parthenogenesis, (3) male killing and (4) cytoplasmic incompatibility. These mechanisms lead to an accelerated spread of *Wolbachia* through populations. As the infection is transmitted exclusively maternally, positive selection acts on the infected females mitochondrial DNA, which results into selective sweeps and a homogenization of mitochondrial haplotypes across the population. *Wolbachia*- induced selective sweeps can spread through closely related populations or species via hybridization events, leading to misinterpretation of data based on mitochondrial DNA (Jiggins 2003). Interestingly, the influence of bacterial reproductive parasites, e.g., *Wolbachia*, on DNA barcoding remains controversial. However, scenarios resulting in the misinterpretation of species diversity are well supported (Gerth *et al.* 2011). Stalhut *et al.* (2012) correctly note that neither effects on host mitochondrial DNA (mtDNA) nor stable infections can be detected by PCR-based *Wolbachia* screens of single individuals. However, most prevalence distributions of *Wolbachia* within host species follow a most-or-few-pattern and by testing only a few individuals per species, the actual prevalence is likely to be underestimated rather than overestimated (Hilgenboecker *et al.* 2008; Zug & Hammerstein 2012). Even more importantly, ancient *Wolbachia* infections that are not traceable today may still lead to biased conclusions drawn from mtDNA. Whether or not the presence of such endosymbionts has at any time resulted in non-neutral inheritance of mitochondria cannot be concluded from the presence or absence of these alone. Nonetheless, a straightforward test of the

performance of mitochondrial markers such as *cox1* can be conducted by co-analysing nuclear markers. Given the growing body of evidence for potentially *Wolbachia*-induced effects on patterns of mtDNA variation in populations (Bachtrog *et al.* 2006; Narita *et al.* 2006; Whitworth *et al*. 2007; Gompert *et al*. 2008; Charlat *et al*. 2009; Nice *et al*. 2009; Raychoudhury *et al*. 2009; Atyame *et al*. 2011; Dyer *et al.* 2011; Graham & Wilson 2012; Xiao *et al.* 2012), we think it is premature to deem these negligible for DNA barcoding. This problem has been repeatedly discussed in the literature; however, most barcoding studies dealing with terrestrial arthropods hardly address it (e.g., Sheffield *et al.* 2009; Smith *et al.* 2012).

Stalhut *et al.* (2012) describe DNA barcoding as a tool that needs close scrutiny and taxonomic background knowledge to be interpreted accurately, analogous to taxonomic keys. As such, barcode libraries can indeed be valuable and help to identify research questions. However, in the DNA barcoding literature, "democratizing taxonomy" is repeatedly being praised as achievement and a main goal of DNA barcoding (Hebert & Gregory 2005; Holloway 2006; Janzen *et al.* 2009). The availability and applicability of barcodes for everyone, especially non-experts, was regarded as a big advantage of the DNA barcodes (in contrast to, e.g., taxonomic keys, Packer *et al.* 2009). If on the other hand users of barcode libraries do need expert knowledge anyway, a simplistic one-marker approach seems not adequate. Furthermore, although Stalhut *et al.* (2012) claim that DNA barcodes are not used with disregard to other evidence, DNA barcoding studies relying on a single marker are numerous (Sheffield *et al.* 2009; Zaldívar-Riverón *et al.* 2010 reviewed in Taylor & Harris 2012; e.g., Hausmann *et al.* 2011; Magnacca & Brown 2012).

In our opinion, a multi locus sequence typing (MLST) approach could largely diminish the problems that are associated with DNA barcoding as it is currently practiced, inherited endosymbionts being only one issue among many (Galtier *et al.* 2009). Such MLST-approaches are already widely used for the identification of prokaryotes, as for example *Wolbachia* strains themselves (Baldo *et al.* 2006). Conveniently, a primer tool box which enables the amplification of a variety of markers containing both exonic and fast-evolving intronic nucleotides was recently published for Hymenoptera (Hartig *et al.* 2012). Especially barcoding projects dealing with groups lacking a fundamental taxonomy (e.g., many tropical bee communities) would greatly benefit from using a MLST approach. Consequently, we strongly advocate the inclusion of additional nuclear markers in future barcoding endeavours.

References

Atyame CM, Delsuc F, Pasteur N, Weill M, Duron O (2011) Diversification of *Wolbachia* endosymbiont in the *Culex pipiens* mosquito. *Molecular Biology and Evolution,* **28,** 2761-2772.

Bachtrog D, Thornton K, Clark A, Andolfatto P (2006) Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution,* **60,** 292-302.

Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Applied and Environmental Microbiology,* **72,** 7098-7110.

Charlat S, Duplouy A, Hornett EA, Dyson EA, Davies N, Roderick GK, Wedell N, Hurst GD (2009) The joint evolutionary histories of Wolbachia and mitochondria in *Hypolimnas bolina*. *BMC Evolutionary Biology,* **9,** 64.

Dyer KA, Burke C, Jaenike J (2011) *Wolbachia*-mediated persistence of mtDNA from a potentially extinct species. *Molecular Ecology,* **20,** 2805-2817.

Galtier N, Nabholz B, Glemin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology,* **18,** 4541-4550.

Gerth M, Geißler A, Bleidorn C (2011) Wolbachia infections in bees (Anthophila) and possible implications for DNA barcoding. *Systematics and Biodiversity,* **9,** 319-327.

Gompert Z, Forister ML, Fordyce JA, Nice CC (2008) Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in Lycaeides. *Molecular Ecology,* **17,** 5231-5244.

Graham R, Wilson K (2012) Male-killing *Wolbachia* and mitochondrial selective sweep in a migratory African insect. *BMC Evolutionary Biology,* **12,** 204.

Hartig G, Peters RS, Borner J, Etzbauer C, Misof B, Niehuis O (2012) Oligonucleotide Primers for Targeted Amplification of Single-Copy Nuclear Genes in Apocritan Hymenoptera. *PLoS ONE,* **7,** e39826.

Hausmann A, Haszprunar G, Segerer AH, Speidel W, Behounek G, Hebert PDN (2011) Now DNA-barcoded: the butterflies and larger moths of Germany. *Spixiana (Supplement),* **34,** 47-58.

Hebert PDN, Ratnasingham S, de Waard JR (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B*, **270**, Suppl 1, S96-S99.

Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. *Systematic Biology,* **54,** 852-859.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH (2008) How many species are infected with *Wolbachia*? A statistical analysis of current data. *FEMS Microbiology Letters,* **281,** 215-220.

Holloway M (2006) Democratizing Taxonomy. *Conservation in Practice,* **7**.

Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J-M, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID, Harvey DJ, Hausmann A, Kitching IJ, Lafontaine DON, Landry J-F, Lemaire C, Miller JY, Miller JS, Miller LEE, Miller SE, Montero J, Munroe E, Green SR, Ratnasingham S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources,* **9,** 1-26.

Jiggins FM (2003) Male-killing Wolbachia and mitochondrial DNA: Selective sweeps, hybrid introgression and parasite population dynamics. *Genetics,* **164,** 5-12.

Magnacca KN, Brown MJF (2012) DNA barcoding a regional fauna: Irish solitary bees. *Molecular Ecology Resources*, **12**, 990-998.

Narita S, Nomura M, Kato Y, Fukatsu T (2006) Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. *Molecular Ecology,* **15,** 1095-1108.

Nice CC, Gompert Z, Forister ML, Fordyce JA (2009) An unseen foe in arthropod conservation efforts: The case of *Wolbachia* infections in the Karner blue butterfly. *Biological Conservation,* **142,** 3137-3146.

Packer L, Gibbs J, Sheffield C, Hanner R (2009) DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources,* **9,** 42-50.

Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH (2009) Modes of Acquisition Of Wolbachia: Horizontal Transfer, Hybrid Introgression, and Codivergence in the Nasonia Species Complex. *Evolution,* **63,** 165-183.

Sheffield CS, Hebert PDN, Kevan PG, Packer L (2009) DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. *Molecular Ecology Resources,* **9,** 196-207.

Smith MA, Bertrand C, Crosby K, Eveleigh ES, Fernandez-Triana J, Fisher BL, Gibbs J, Hajibabaei M, Hallwachs W, Hind K, Hrcek J, Huang D-W, Janda M, Janzen DH, Li Y, Miller SE, Packer L, Quicke D, Ratnasingham S, Rodriguez J, Rougerie R, Shaw MR, Sheffield C, Stahlhut JK, Steinke D, Whitfield J, Wood M, Zhou X (2012) *Wolbachia* and DNA Barcoding Insects: Patterns, Potential, and Problems. *PLoS ONE,* **7,** e36514.

Stahlhut JK, Gibbs J, Sheffield CS, Alex Smith M, Packer L (2012) *Wolbachia* (Rickettsiales) infections and bee (Apoidea) barcoding: a response to Gerth et al. *Systematics and Biodiversity,* **10,** 395-401.

Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, **12**, 377-388.

Whitworth TL, Dawson RD, Magalon H, Baudry E (2007) DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B,* **274**, 1731-1739.

Xiao J-H, Wang N-X, Murphy RW, Cook J, Jia L-Y, Huang D-W (2012) *Wolbachia* infection and dramatic intraspecific mitochondrial DNA divergence in a fig wasp. *Evolution,* **66**, 1907-1916.

Zaldívar-Riverón A, Martínez JJ, Ceccarelli FS, De Jesús-Bonilla VS, Rodríguez-Pérez AC, Reséndiz-Flores A, Smith MA (2010) DNA barcoding a highly diverse group of parasitoid wasps (Braconidae: Doryctinae) from a Mexican nature reserve. *Mitochondrial DNA,* **21**, 18-23.

Zug R, Hammerstein P (2012) Still a Host of Hosts for *Wolbachia*: Analysis of Recent Data Suggests That 40% of Terrestrial Arthropod Species Are Infected. *PLoS ONE,* **7**, e38544.

Ancient DNA – retrospective sequencing of types

– Michael Hofreiter, Chenhong Li, Gavin Naylor, Tanja Haus and Kirsten Richter
University of York, Department of Biology, Area 2, York, United Kingdom

University of Potsdam, Institute of Biochemistry and Biology, Department Evolutionary Adaptive Genomics, Potsdam, Germany

DNA sequencing – both in the form of short “barcoding” sequences as well as in the form of complete genomes – has revolutionised both evolutionary and taxonomic research. However, despite the many insights provided by DNA sequencing it has one major drawback: it usually requires a certain amount of tissue to be removed from the specimens in order to obtain sequence data. This is often a trivial requirement, but it can become problematic if type specimens are affected, as morphological integrity is of key importance for type specimens. In addition to this issue, type specimens are often problematic for DNA analyses as they have often been sampled decades or longer ago and were usually prepared and stored in ways that are highly detrimental to DNA survival, resulting in low amounts of low-quality DNA that can be retrieved from such samples. Thus, taxonomic research is in the paradoxical situation that one of its most powerful tools – DNA sequencing – becomes a blunt weapon when it comes to the samples that are at the core of taxonomy, the type specimens.
In my presentation I will show with examples ranging from golden moles to river sharks how the various problems can be overcome, making type specimens potentially accessible for taxonomic research without – or at least restricting – damage to their morphological integrity. The tools that have become available in recent years include non-destructive DNA extraction methods as well as sensitive PCR amplifications, next generation sequencing and DNA hybridisation capture. This tool box has incredible potential for DNA analyses of type specimens, but I will also show that there is still a lot of scope for methodological improvements.

Next generation DNA sequencing methods now permit the genetic delineation of extremely young (cichlid fish) species and populations

– Axel Meyer

University Konstanz, Faculty of Biology, Department Zoology/Evolutionary Biology, Konstanz, Germany

Cichlid fishes are well-known for their extreme species-richness and for their record rates of speciation. In Lake Victoria in East Africa more than 500 endemic species evolved in less than 100,000 years and several Nicaraguan crater lakes, some of which are less than 2,000 years old, yet contain several extremely young endemic species. These attributes make cichlids a favourite group of organisms for the study of speciation, yet, so far it has been impossible to resolve the phylogenetic relationships or to even find autoapomorphic molecular markers to designate species reliably. The standard DNA-Barcodes all fail since all cichlids from Lake Victoria, for example, are identical in their COI and cytochrome b sequences. Moreover, phenomena such as incomplete lineage sorting and hybridization add to the difficulty to sorting and delineating evolutionary lineages. New Next-generation-Sequencing (NGS) techniques now provide massive amounts of genomic information, that, for the first time, permit to molecularly differentiate species and to reconstruct their phylogenetic relationships.

Theses

- Massive amounts of NGS will permit to study the evolutionary history and the phylogenetic relationships among very young species, that so far, eluded this type of question.
- Which NGS methods, data sets, and analyses will prove to be most useful in this regard?

1KITE – 1K Insect Transcriptome Evolution

– Bernhard Misof

Zoological Research Museum Alexander Koenig – Leibniz-Institute for Animal Biodiversity, University Bonn, Molecular Biodiversity Research, Bonn, Germany

Transcriptomics is developing with an extreme pace. It enables scientists to study gene expression patterns much more efficiently than previously achieved and thus delivers the toolbox to study the evolution of genetic pathways and eventually phenotypes. Additionally, transcriptomics is becoming the favorite approach in phylogenomics, because it can deliver an unprecedented set of informative gene markers for phylogenetic studies. It harbors thus also the potential to contribute to taxonomic research based on multi-gene approaches. In taxonomy multi-gene approaches have not been seriously explored because of a lack of sufficiently good data to design new PCR targets. Additionally, multi-gene approaches seem to be less-time efficient, too expensive and therefore essentially untractable.
In my presentation, I will analyse whether transcriptomics can be of any use for taxonomic research. I will particularly focus on the questions, whether

(1) Transcriptomics can deliver exhaustive sets of 1:1 orthologs for an arbitrary taxonomic set of organisms
(2) Transcriptomics could be a feasible standard approach in species identification
(3) Transcriptomics can be used to develop operational metagenomic tools to be used in Metazoan.

Additionally, I will explore the differences in the quality of genomic and transcriptomic data currently produced in the international community and will present some of the perspectives developing from these international activities.

The 1KITE consortium (www.1kite.org) represents an international group of 60 scientist with the mission of sequencing 1000 insect transcriptomes covering the taxonomic breath of insects and delivering a robust backbone tree of this megadiverse group of metazoan organisms until the end of 2014. The Bejing Genomics Institute (BGI) is financing this project with 6Million US dollars. In order to achieve this goal the 1KITE scientists have to explore whether methods and tools currently used in phylogenetics can be scaled to these data dimensions and if not have to scope with this problem in developing or adapting new approaches.
Within the 1KITE consortium, we managed to have >800 species either sequenced or even fully assembled and analysed by now, Dec. 2012. Given the pace of collecting, we can be confident to finish sequencing of these 1000 transcriptomes in 2012/2013.
It has been agreed upon with the 1KITE consortium to use paired-and sequencing of 250 bp insert libraries on a HiSeq2000 platform. We sequence 2.5 G bases of raw nucleotide sequence data for each specimen. In order to improve the assembly, we have developed in cooperation with the BGI a dedicated transcriptome read assembler SOAPdenovo-trans, which is publicly available. We have further adapated and extended a published approach (HAMSTR, Ebersberger et al. 2009) to assign 1:1 orthologous for each transcriptome. Starting from a reference set of orthologous genes derived from full genome sequences we are able to identify almost the full set of these genes (99%) in each transcriptome. This amounts e.g. to 1478 genes across all hexapods or 4.500 genes for all Hymenop-

tera. We are thus identifying the full potentially comparable and amplifiable genes across a taxonomic group of interest and deliver comparable data quality to full genome sequencing.
Turning to genomics, I will shortly present our description of a new strepsipteran species including the sequencing of the full genome jointly published with scientists from the University of Jena and Münster. Additionally, I will introduce the i5k initiative which proposed to sequence 5000 insect genomes within the next five years. My group is part of the core unit of i5k and is sequencing genomes of 9 representatives of various insect orders in cooperation with the Baylor Institute of Medicine, USA.
Taking the success of transcriptome and genome sequencing together, we can see that to delimit species the design of multi-gene data using the targeted PCR approach is feasible. However, a full transcriptome or genome sequencing for each new species is currently still out of scope mostly due to financial and most importantly, analytical and computational reasons. This situation could change soon with the development of the third generation sequencing techniques.
We have demonstrated the success of the proposed design of multi-gene targeted PCR approaches by developing a primer toolbox for Hymenoptera.
The extensive transcriptomic/genomic data can be mined for many speciation relevant genes and provides thus an extremely rich source of data for taxonomic research.

Sources:
1KITE consortium:www.1kite.org
i5k initiative:www.arthropodgenomes.org

References

Ebersberger I, Strauss S, von Haeseler A (2009). HaMStR: Profile hidden markov model based search for orthologs in ESTs. *BMC Evolutionary Biology*, **9**, 157.

Hartig G, Peters RS, Borner J, Etzbauer C, Misof B, Niehuis O (2012). Oligonucleotide primers for targeted amplification of single-copy nuclear genes in apocritan Hymenoptera. *PLoS ONE,* **7** (6), e39826. doi:10.1371/journal.pone.0039826

Meusemann K, von Reumont BM, Simon S, Roeding F, Kueck P, Ebersberger I, Strauss S, Walzl M, Pass G, Breuers S, Achter V, Wägele J-W, Hadrys H, Burmester T, von Haeseler A, Misof B (2010). A phylogenomic approach to resolve the arthropod tree of life.*Mol. Biol. Evol.,* **27** (11), 2451-2464.

Niehuis O, Hartig G, Grath S, Pohl H, Lehmann J, Tafer H, Donath A, Krauss V, Eisenhardt C, Hertel J, Petersen M, Mayer C, Meusemann K, Peters RS, Stadler PF, Beutel RG, Bornberg-Bauer E, McKenna DD, Misof B (2012). Genomic and morphological evidence converge to resolve the enigma of Strepsiptera. *Current Biology***22**, 1309-1313.

Pohl H, Niehuis O, Gloyna K, Misof B, Beutel R (2012). A new species of *Mengenilla* (Insecta, Strepsiptera) from Tunisia. *ZooKeys,* **198**, 79-101.

Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, Goldsmith MR, Lawson D, Okamuro J, Robertson HM, Schneider DJ (2011). Creating a buzz about insect genomes. *Science*, **331**, 1386.

The use of DNA sequence data in animal identification and classification

– Michael J. Raupach

Senckenberg am Meer, German Center for Marine Biodiversity Research (DZMB), Department Molecular Taxonomy, Wilhelmshaven, Germany

Correct species identification and classification represents a fundamental aspect in most biological sciences (Figure 1). Nevertheless, a routine determination of many species is difficult, time-consuming, requires highly specialized knowledge, and the identification of larval stages or fragments of organisms using conventional morphological methods constitutes an impossible task for many taxa. Therefore, correct species identification represents a limiting factor in most biodiversity assessments and ecological studies (e.g. Floyd *et al*. 2002, Smith *et al*. 2005, Hajibabaei *et al*. 2007). In this context, the use of DNA sequence data represents a promising and effective tool for fast and accurate species identification (see Janzen *et al*. 2009, Bucklin *et al*. 2011, Nagy *et al*. 2012) and classification (Tautz *et al*. 2003, Vogler & Monagham 2007).

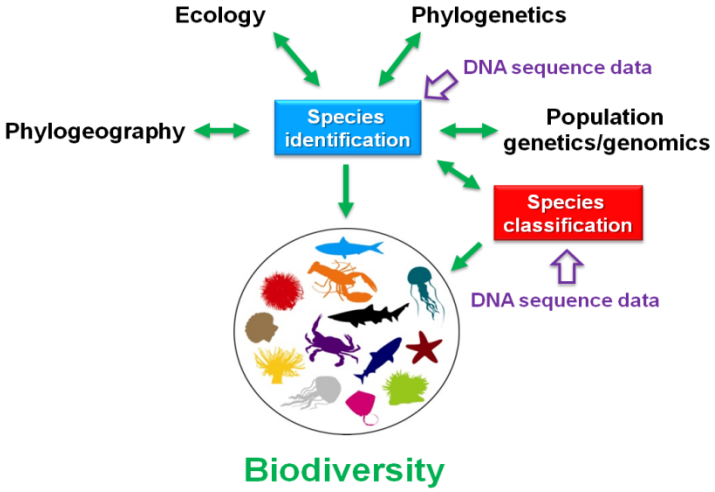

Figure 1: DNA sequence data and other molecular data can help to accelerate species identification and classification.

Species identification using DNA sequences: DNA barcoding

For animals, mitochondrial DNA exhibits several characteristics that makes it highly attractive for molecular identification, namely the generally high substitution rates, the almost exclusively maternal inheritance, and the lack of recombination (Ballard & Whitlock 2004, Ballard & Rand 2005). Moreover, because of uniparental inheritance and haploidy, mtDNA has a four-fold smaller effective population size compared to nuclear DNA, leading to faster lineage sorting (Birky *et al*. 1983). Some years ago, an app. 650 base-pair fragment of the mitochondrial cytochrome *c* oxidase I (COI) gene was proposed as global standard, the so-called “barcode region” for animals (Hebert *et al*. 2003). The concept of DNA barcoding relies on low levels of mtDNA variation within species in combination with clear genetic differentiation between species. However, the exclusive use of COI or mitochondrial DNA in general is not without risks:

- DNA barcoding can overestimate the number of species when nuclear mitochondrial pseudogenes (numts) are co-amplified, as it has been demonstrated for some taxa (e.g. An-

tunes *et al*. 2007, Song *et al*. 2008). Here, the use of quality control measures, e.g. sequence checking for non-synonymous mutations, premature stop codons or insertion-deletions, can help to detect numts in most cases (Song *et al*. 2008).

- Heteroplasmy, the presence of a mixture of more than one type of an organelle genome (e.g. mtDNA) within a cell or individual, is typically considered as a transitional and short-lived state of metazoan mitochondrial evolution, but can confuse the identification system also (Doublet *et al*. 2008). However, heteroplasmy is rare and known only for some very few taxa, e.g. the genus *Mytilus* (Skibinski *et al*. 1994, Quesada *et al*. 1996) or the bee genus *Hylaeus* (Magnacca & Brown 2012).
- DNA barcoding will fail if species hybridize or when species pairs have very recent origins. For some time after the initial split, new sister species will share alleles, either because of ongoing gene flow, or because of recent ancestry. In such cases, sequences from one or few individuals will not be sufficient for an unequivocal assignment to a particular group (Tautz *et al*. 2003).
- Maternally inherited endosymbionts such as the Alpha-proteobacterium *Wolbachia* may cause a linkage disequilibrium within mtDNA of terrestrial arthropods, resulting in a homogenization of mtDNA haplotypes (e.g. Jiggins *et al*. 2001, Duron *et al*. 2008). Further studies are needed to analyze factual effects of these inherited parasites on mtDNA variability on a broader scale more in detail.

However, numerous studies show that the variability of mitochondrial DNA of many species is predominately driven by phylogeographic events including population expansions, population bottlenecks, vicariance and migration as consequence of historical processes (Avise 2009, Hickerson *et al*. 2010). For example, the genetic structure and distribution of the current European biota has been significantly shaped by Pleistocene glacial oscillations (Hewitt 1996, 2000, 2004). While high latitudes of Europe were covered with ice and permafrost, temperate regions were compressed to the South. When temperatures were at their coldest, populations of temperate-adapted species often persisted in geographically isolated ice-free refugia, where the climate was less inhospitable (Taberlet *et al*. 1998, Hewitt 1999, 2004, Schmitt 2007). As the climate warmed and glaciers retreated, northern Europe was recolonized from these refugia by leading-edge expansion of populations (Taberlet *et al*. 1998, Hewitt 1999, 2004). Not surprisingly, the impact of such historical processes on the genetic variability of mtDNA differs from species to species. Depending on the degree of variability, the usefulness of COI for correct species identification is affected also.

In spite of the already mentioned potential problems, DNA barcoding has been highly successfully applied in an amazing vast number of taxa, both aquatic and terrestrial, for species identification (e.g. Costa *et al*. 2007, Zhou *et al*. 2009, Raupach *et al*. 2010, Lakra *et al*. 2011, Lijtmaer *et al*. 2011, Spelda *et al*. 2011). The use of DNA barcodes also helped in the discovery of numerous cryptic or overseen species (e.g. Hebert *et al*. 2004, Bradford *et al*. 2009). Consequently, many new species descriptions include barcode data (e.g. Fisher & Smith 2008, Hendrich & Balke 2011, Butcher *et al*. 2012, Riehl & Kaiser 2012).

Classification of organisms based on DNA sequences: DNA taxonomy

Beside the identification of known species, the delimitation as well as classification of unknown species using molecular data is highly desirable. Here, so-called coalescent-based species delimitation approaches are frequently used (Fujita *et al*. 2012). The central aim of coalescent-based approaches

is to identify independently evolving lineages as a transition from coalescent to speciation branching patterns on a phylogenetic tree, with each lineage representing individually species. Using mitochondrial data, however, former bottleneck events or selective sweeps can become problematic in reconstructing the coalescence of mtDNA lineages and therefore for species delineation. It should be also kept in mind that such methods are also sensitive to introgression and incomplete lineage sorting.

While various studies reveal the potential of this method (e.g. Pons *et al*. 2006, Papadopoulou *et al*. 2009), it is clear that molecular species classification cannot be based on a standardized single marker system in many cases (Fujita *et al*. 2012). Recent studies clearly favor a multi-locus integrative approach, especially for closely related species (Yang & Rannala 2010, Fujita *et al*. 2012). In contrast to single-gene datasets, multilocus data can test alternative hypotheses of lineage divergence that allow discordance for gene trees under genetic drift.

Summary and Perspective: The future of taxonomy

As consequence of the tremendous advances in molecular biology, new methods have become available to accelerate the identification and classification of organisms. In this context, DNA barcoding and DNA taxonomy represent two highly valuable concepts and methods in modern biodiversity assessment studies. Based on massive transcriptome and/or genome analysis using next-generation sequencing, it is obvious that new promising markers will become identified for currently problematic taxa in the near future.

In addition to sequence data, other methods may play an important role in species identification and classification also, e.g. the use of proteome data generated with matrix-assisted laser desorption/ionization time-of-flight spectrometry (MALDI TOF), which has been established in microbiology for years (e.g. Marvin *et al*. 2003, Degand *et al*. 2008, Wang *et al*. 2012). Pioneering studies analyzing eukaryotic organisms reveal the high potential of this approach (e.g. Mazzeo *et al*. 2008, Riccardi *et al*. 2012). Other promising approaches are Raman (e.g. Ashton *et al*. 2011) or near IR spectroscopy (e.g. Fischnaller *et al*. 2012).

All shown methods and approaches make clear that taxonomy has to adapt new methods: not only organisms evolve, taxonomy must also. In terms of modern and efficient taxonomy, new species descriptions without sequence or proteome data are incomplete.

Literature

Antunes A, Pontius J, Ramos MJ, O´Brien SJ, Johnson WE (2007). Mitochondrial introgressions into the nuclear genome of the domestic cat. *Journal of Heredity*, **98**, 414-420.

Ashton L, Lau K, Winder CL, Goodacre R (2011). Raman spectroscopy: lighting up the future of microbial identification. *Future Microbiology*, **6**, 991-997.

Avise JC (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3-15.

Ballard JWO, Whitlock MC (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729-744.

Ballard JWO, Rand DM (2005). The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Reviews in Ecology and Systematics*, **36**, 621-642.

Birky Jr CW, Maruyama T, Fuerst PA (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**, 513-527.

Bradford T, Adams M, Humphreys WF, Austin AD, Cooper SJB (2009). DNA barcoding of stygofauna uncovers cryptic amphipod diversity in a calcrete aquifer in Western Australia`s arid zone. *Molecular Ecology Resources*, **10**, 41-50.

Bucklin A, Steinke D, Blanco-Bercial L (2011). DNA barcoding of marine Metazoa. *Annual Review of Marine Science*, **3**, 471-508.

Butcher BA, Smith MA, Sharkey MJ, Quicke DLJ (2012). A turbo-taxonomic study of Thai *Aleiodes*(*Aleiodes*) and *Aleiodes* (*Arcaleiodes*) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. *Zootaxa*, **3457**, 1-232.

Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN (2007). Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Science*, **64**, 272-295.

Degand N, Carbonelle E, Dauphin B, Beretti J-L, Le Bourgeois M, Sermet-Gaudelus I, Segonds C, Berche P, Nassif X, Ferroni A (2008). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of nonfermenting gram-negative bacilli isolated from cystic fibrosis patients. *Journal of Clinical Microbiology*, **46**, 3361-3367.

Doublet V, Souty-Grosset C, Bouchon D, Cordaux R, Marcade I (2008). A thirty million year-old inherited heteroplasmy. *PLoS ONE*, **3**, e2938.

Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, Hurst GD (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biology*, **6**, 27.

Fischnaller S, Dowell FE, Lusser A, Schlick-Steiner BC, Steiner FM (2012). Non-destructive species identification of *Drosophila obscura* and *Drosophila subobscura* (Diptera) using near-infrared spectroscopy. *Fly*, **6**, 284-289.

Fisher BL, Smith MA (2008). A revision of malagasy species of*Anochetus* Mayr and *Odontomachus* Latreille (Hymenoptera: Formicidae). *PLoS Biology*, **3**, e1787.

Floyd R, Abebe E, Papert A, Blaxter M (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology*, **11**, 839-850.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, **27**, 480-488.

Hajibabaei M, Singer GAC, Clare EL, Hebert PDN (2007). Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology*, **5**, 24.

Hebert PDN, Ratnasingham S, de Waard JR (2003). Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **270**, S96-S99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly*Astrapes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812-14817.

Hendrich L, Balke M (2011). A simultaneous journal/wiki publication and dissemination of a new species description: *Neobidessodes darwiniensis* sp. n. from northern Australia (Coleoptera, Dytiscidae, Bidessini). *Zookeys*, **79**, 11-20.

Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hewitt GM (1999). Post-glacial re-colonisation of European biota. *Biological Journal of the Linnean Society*, **68**, 87-112.

Hewitt GM (2000). The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hewitt GM (2004). The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology*, **1**, 4.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010). Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291-301.

Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J-M, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID, Harvey DJ, Hausmann A, Kitching IJ, Lafontaine D, Landry J-F, Lemaire C, Miller JY, Miller JS, Miller L, Miller SE, Montero J, Munroe E, Green SR, Ratnasingham S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ (2009). Integration of

DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources*, **9**, 1-26.

Jiggins FM, Hurst GDD, von der Schulenburg JHG, Majerus MEN (2001). Two malekilling *Wolbachia* strains coexist within a population of the butterfly *Acreae encedon*. *Heredity*, **86**, 161-166.

Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, Punia P, Gopalakrishan A, Singh KV, Ward RD, Hebert PDN (2011). DNA barcoding Indian marine fishes. *Molecular Ecology Resources*, **11**, 60-71.

Lijtmaer DA, Kerr KCR, Barreira AS, Hebert PDN, Tubaro PL (2011). DNA barcode libraries provide insight into continental patterns of avian diversification. *PLoS ONE*, **6**, e20744.

Magnacca KN, Brown MJF (2010). Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). *BMC Evolutionary* Biology, **10**, 174.

Marvin LF, Roberts MA, Fay LB (2003). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clinica Chimica Acta*, **337**, 11-21.

Mazzeo MF, de Giulio B, Guerriero G, Ciarcia G, Malorni A, Russo GL, Siciliano RA (2008). Fish authentication by MALDI-TOF mass spectrometry. *Journal of Agricultural and Food Chemistry*, **56**, 11071-11076.

Nagy ZT, Sonet G, Glaw F, Vences M (2012). First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. *PLoS ONE*, **7**, e34506.

Papadopoulou A, Jones AG, Hammond PM, Vogler AP (2009). DNA taxonomy and phylogeography of beetles of the Falkland Islands (Islas Malvinas). *Molecular Phylogenetics and Evolution*, **53**, 935-947.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595-609.

Quesada H, Skibinski DA, Skibinski DO (1996).Sex-biased heteroplasmy and mitochondrial DNA inheritance in the mussel *Mytilus galloprovincialis* Lmk. *Current Genetics*, **29**, 423-426.

Raupach MJ, Astrin JJ, Hannig K, Peters MK, Stoeckle MY, Wägele JW (2010). Molecular identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology*, **7**, 26.

Riccardi N, Lucini L, Benagli C, Welker M, Wicht B, Tonolla M (2012). Potential of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for the identification of freshwater zooplankton: a pilot study with three *Eudiaptomus* (Copepoda: Diaptomidae) species. *Journal of Plankton Research*, **34**, 1-9.

Riehl T, Kaiser S (2012). Conquered from the deep sea? A new deep-sea isopod species from the Antarctic shelf shows pattern of recent colonization. *PLoS ONE*, **7**, e49354.

Schmitt T (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Skibinski DOF, Gallagher C, Beynon CM (1994). Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. Genetics 138: 801-809.

Smith MA, Fisher BL, Hebert PDN (2005). DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society of London Series B*, **360**, 1825-1834.

Song H, Buhay JE, Whiting MF, Crandall KA (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13486-13491.

Spelda J, Reip HS, Oliveira-Biener U, Melzer RR (2011). Barcoding Fauna Bavarica: Myriapoda - a contribution to DNA sequence-based identifications of centipedes and millipedes (Chilopoda, Diplopoda). *ZooKeys*, **156**, 123-139.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998). Comparative phylogeography and postglacial colonisation routes in Europe. *Molecular Ecology*, **7**, 453-464.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003). A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70-74.

Vogler AP, Monagham MT (2007). Recent advances in DNA taxonomy. *Journal of Systematic and Evolutionary Research*, **45**, 1-10.

Wang Y, Zhou Q, Li B, Liu B, Wu G, Ibrahim M, Xie G, Li H, Sun G (2012). Differentiation in MALDI-TOF MS and FTIR spectra between two closely related species *Acidovorax oryzae* and *Acidovorax citrulli*. *BMC Microbiology*, **12**, 182.

Yang Z, Rannala B (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9264-9269.

Zhou X, Adamowicz SJ, Jacobus LM, DeWalt RE, Hebert PDN (2009). Towards a comprehensive barcode library for arctic life-Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology*, **6**, 30.

Natural hybridization in primates and what "-omics" contributed to primate taxonomy and systematics

– Christian Roos, Dietmar Zinner

German Primate Center, Gene Bank of Primates, Primate Genetics Laboratory, Göttingen, Germany

„Nothing in biology makes sense except in the light of evolution" (Theodosius Dobzhansky 1973)

„Without taxonomy to give shape to the bricks, and systematics to tell us how to put them together, the house of biological science is a meaningless jumble" (Robert M. May 1990)

These two often cited sentences clearly reflect the fact that the modern evolutionary synthesis is the overall basis of biology. Both, taxonomy (the science of identifying, naming and classifying organisms) and systematics (the science of reconstructing phylogenetic relationships), are essential for evolutionary and biodiversity research (Edwards & Cavalli-Sforza 1964; Harvey et al. 1996; Wiley et al. 1991) as well as for many other biological disciplines. This includes the determination of evolutionary units (EUs), which can be identified and diagnosed due to their distinct genetic make-up and/or independent demographic history (Avise 2005). This is of great importance when issues related to conservation are tackled, since much of the total genetic diversity within a species is often partitioned among sets of geographic populations (demes) (Avise 2005). Likewise, phylogenetic research is the framework for testing hypotheses concerning the evolution of traits, e.g. physiology, morphology, behaviour, diseases, or co-evolutionary processes. In particular, for comparative analyses of adaptive processes, and for the discrimination between ancestral and derived states, taxonomy and systematics, i.e. a robust phylogeny are inevitable (Bjork et al. 2011; Rohland et al. 2007; Shultz et al. 2011; Switzer et al. 2005; Zinner et al. 2009, 2011). Consequently, any change in a specific character can be tracked back into the past (Losos 2011).

Non-human primates are our closest living relatives and due to their similar anatomy, physiology and genetic make-up, various non-human primate species are used as model organisms in biomedical research, in particular in preclinical studies, virology, toxicology and neuroscience. Similarly, non-human primates are appropriate model species to study the roots of human behaviour and sociality, and to trace back and understand human evolution and genomics (Enard & Pääbo 2004; Goodman et al. 2005; Strum 2012). Finally, nearly half of all primate species are threatened and some are even close to extinction (IUCN 2012), so that immediate conservation actions are required. For all these biological and medical disciplines and for conservation purposes, knowledge about the taxonomic/genetic diversity of primates and their evolutionary history is an essential cornerstone.

For example, rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*), the two most commonly used non-human primate model species in biomedical research, have wide distributions in Asia. They exhibit extreme intra-specific genetic diversity, which most likely impact susceptibility and resistance to various diseases. Actually, local populations of rhesus macaques carry different MHC alleles, which are known to affect the survival time of experimentally SIV infected animals (Sauermann et al. 2008) and accordingly influence SIV experiments. Consequently, genome-wide markers are urgently required to select appropriate individuals for respective experiments.

Likewise, the question: “What makes us human?” can only be answered by comparative analyses of data derived from the genome and transcriptome of humans and chimpanzees. Humans and chimpanzees differ in less than 2% of their genomes, but analyses on RNA level have shown that both species show species-specific gene expression patterns, with main differences particularly pronounced in the brain (Enard et al. 2002).

Molecular work similarly contributed to a better understanding of primate diversity. In the last two decades, the number of primate species increased dramatically from 230 in 1996 (Rowe 1996) to 480 in 2013 (Mittermeier & Wilson 2013), primarily as a result of applying the Phylogenetic Species Concept (Cracraft 1983) instead of the Biological Species Concept (Dobzhansky 1937; Mayr 1942), but also due to the discovery of numerous cryptic species by applying molecular methods. This is particularly true for nocturnal primates, because morphological differences are small and often not diagnostic, and only molecular studies revealed that these taxa are more diverse than previously believed. Likewise, molecular studies expanded our knowledge about distributional ranges of taxa, allowing better and more efficient conservation measurements.

Surprisingly, in course of many molecular studies on primate diversity and phylogeny, numerous cases of natural hybridization were uncovered. In fact, it is now estimated that natural hybridization occurs in more than 10% of all primate species and it was observed in all major radiations (Zinner et al. 2011). Hybridization, as the admixture of previously isolated gene pools, was not expected at such a magnitude, because as in other animal species, inter-specific hybridization was regarded as an evolutionary dead end, since hybrid offspring often show limited vitality or were sterile, as for instance in mules (horse x donkey). In some cases, however, the exchange of genes between species may have accelerated adaptation and may have led to the formation of new lineages. Thus, hybridization can be regarded as one important evolutionary mechanism driving speciation processes. Hybridization can be a recent and ongoing process that can be also detected by intermediate morphotypes, e.g., in baboons, or it occurred in the past, which left only traces of genetic exchange in the genome, e.g., in some Asian langur genera. Recent and ancient hybridization events occur/occurred mainly between subspecies and species, but they have also been detected between primate genera and even in the human lineage (Zinner et al. 2011). Besides bidirectional hybridization, where males and females of both parent taxa contribute to hybridization, also unidirectional hybridization can occur, where only one sex of one parent taxon contributes, but not the other. Such an introgression normally results in mitochondrial or Y chromosomal capture, or nuclear swamping, depending on the contributing sex. In any case, one can expect to find indications for mosaic genomes either in the hybrids or the introgressed taxon.

Ancient hybridization events are normally not detectable when analysing morphological traits. Such events are only revealed by molecular studies including various differently inherited molecular markers. Sometimes traces of hybridization in genomes are so minimal that they can only be detected by comparing many different loci and even whole genomes are sometimes required. A prominent example therefore is the gene flow among modern humans, Neanderthals and Denisovans (Figure 1) (Green et al. 2010; Meyer et al. 2012). Respective studies have shown that Neanderthals have contributed ~2% to non-African modern humans and that ~6% of the genomes of Papuans derived from Denisovans, a human species whose fossils have been found in the Denisova Cave in Siberia.

The various cases of natural hybridization events among primate taxa suggest that the Biological Species Concept might be not applicable and outdated at least for primates. Numerous other species

concepts are available, but all of them have their drawbacks. The Phylogenetic Species Concept might be best practical for primates (Groves 2012). However, a general problem in taxonomy remains: the classification of hybrid taxa. Recognizing them all as distinct taxa or at least distinct evolutionary units might be a possible solution. This, however, would increase the number of taxa (species) dramatically.

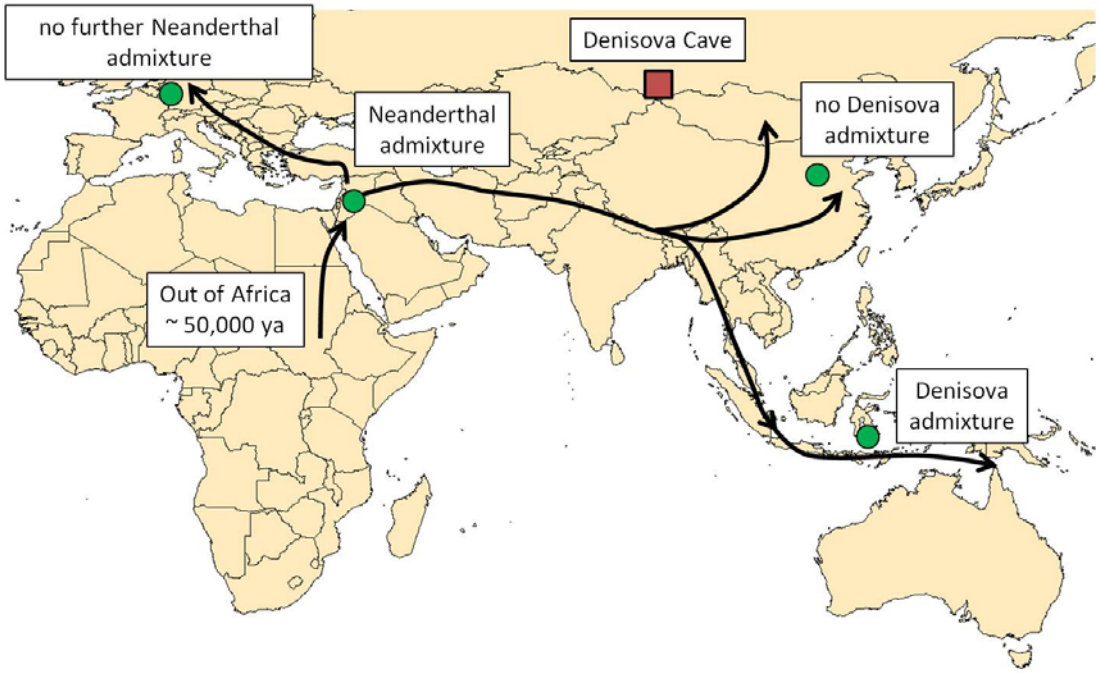

Figure 1: Dispersal of modern humans from Africa.A map illustrating the dispersal of modern humans from Africa about 50,000 years ago, followed by admixture with Neanderthals in the ancestry of all non-Africans, followed by admixture with Denisovans in the ancestry of New Guineans (adapted from Stoneking & Krause 2011).

Main challenges/Relevance to society

1. Research on primate taxonomy and systematics is important for many biological (and medical) research disciplines, because primates are our closest living relatives and accordingly widely used in biomedical research and as model to study human evolution.
2. Only comparative analyses of human and non-human primate genomes, transcriptomes and proteomes will provide the necessary information what makes us human. It will help us to understand the evolution, discovery and interpretation of the genetic underpinnings of human adaptation and diseases.
3. Since Darwin, understanding speciation processes is one of the most important research interests of evolutionary biologists. The many cases of natural hybridization in primates provide good models to test for natural selection and adaptation, and how species evolve.
4. Due to the large intra-specific genetic diversity of biomedical model species, e.g. rhesus and long-tailed macaques, genomic tools are required to select individuals for experiments. This will dramatically reduce the number of individuals in experiments.
5. Many new primate species were discovered in recent years by molecular studies and it can be expected that the full genetic/taxonomic diversity of primates is not fully assessed yet.

6. Nearly 50% of all primate species are threatened and some are even close to extinction, so that immediate conservation actions are required. Genetic methods can help to define distributional ranges and to select appropriate areas for protection.
7. For most if not all these reasons, data from a few genomic loci is not sufficient and complete genome sequences are required, because they allow deep insights into genome evolution and provide the necessary information to address the topics above. The application of high-throughput next-generation sequencing methods will be an efficient, economical and fast way to do so.
8. Analysing the amount of sequence data will be a major challenge and skilled bioinformaticians are required. Taxonomic knowledge is becoming more and more rare. Assign specimens to taxa is the basis of taxonomic research and thus skilled taxonomists are necessary.

References

Avise JC (2005). Phylogenetic units and currencies above and below the species level. In: Purvis A, Brooks T, Gittleman J (eds.), Phylogeny and conservation. Cambridge: Cambridge University Press, 76-100.

Bjork A, Liu W, Wertheim JO, Hahn BH, Worobey M (2011). Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol Biol Evol*, **28**, 615-623.

Cracraft J (1983). Species concepts and speciation analysis. *Curr Ornithol*, **1**, 159-187.

Dobzhansky T (1937). Genetics and the origin of species. New York: Columbia University Press.

Dobzhansky T (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, **35**, 125-129.

Edwards AWF, Cavalli-Sforza LL (1964). Reconstruction of evolutionary trees. In: Heywood VH, McNeill J (eds.), Phenetic and phylogenetic classification. London: Systematics Association, 67-76.

Enard W, Pääbo S (2004). Comparative primate genomics. *Annu Rev Genomics Hum Genet*, **5**, 351-378.

Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R et al. (2002). Intra- and interspecific variation in primate gene expression patterns. *Science*, **296**, 340-343.

Goodman M, Grossmann LI, Wildman DE (2005). Moving primate genomics beyond the chimpanzee genome. *Trends Genet*, **21**, 511-517.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH et al. (2010). A draft sequence of the Neandertal genome. *Science*, **328**, 710-722.

Groves C (2012). Species concepts in primates. *Am J Primatol*, **74**, 687-691.

Harvey PH, Leigh Brown AJ, Maynard Smith J, Nee S (1996). New uses for new phylogenies. Oxford: Oxford University Press.

IUCN (2012). IUCN red list of threatened species 2012.2 [http://www.iucnredlist.org].

Losos JB (2011). Convergence, adaptation, and constraint. *Evolution*, **65**, 1827-1840.

May RM (1990). Taxonomy as destiny. *Nature*, **347**, 129-130.

Mayr E (1942). Systematics and the origin of species. New York: Columbia University Press.

Sauermann U, Siddiqui R, Suh YS, Platzer M, Leuchte N, Meyer H, Mätz-Rensing K, Stoiber H, Nürnberg P, Hunsmann G et al. (2008). MHC class I haplotypes associated with survival time in simian immunodeficiency virus (SIV)-infected rhesus macaques. *Genes Immun*, **9**, 69-80.

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, **338**, 222-226.

Mittermeier RA, Wilson DE (2013). Handbook of the mammals of the world. Volume 3: Primates. Barcelona: Lynx Edicions.

Rohland N, Malaspinas AS, Pollack JL, Slatkin M, Matheus P, Hofreiter M (2007). Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol*, **5**, e207.

Rowe N (1996). The pictorial guide to the living primates. New York: Pogonias Press.

Shultz S, Opie C, Atkinson QD (2011). Stepwise evolution of stable sociality in primates. *Nature*, **479**, 219-222.

Stoneking M, Krause J (2011). Learning about human population history from ancient and modern genomes. *Nature Rev Genet*, **12**, 603-614.

Switzer WM, Salemi M, Shanmugam V, Gao F, Cong ME, Kuiken C, Bhullar V, Beer BE, Vallet D, Gautier-Hion A et al. (2005). Ancient co-speciation of simian foamy viruses and primates. *Nature*, **434**, 376-380.

Strum SC (2012). Darwin's monkey: why baboons can't become human. *Am J Phys Anthropol*, **149** (Suppl 55), 3-23.

Wiley EO, Siegel-Causey DJ, Brooks DR, Funk VA (1991). The compleat cladist: a primer of phylogenetic procedures. *Mus Nat Hist Univ Kansas Spec Publ*, **19**, 1-158.

Zinner D, Groeneveld LF, Keller C, Roos C (2009). Mitochondrial phylogeography of baboons (Papio spp.) - indication for introgressive hybridization? *BMC Evol Biol*, **9**, 83.

Zinner D, Arnold ML, Roos C (2011). The strange blood: natural hybridization in primates. *Evol Anthropol*, **20**, 96-103.

Barcoding microbial eukaryotes

– Thorsten Stoeck
University of Kaiserslautern, Department of Biology, Ecology Group, Kaiserslautern, Germany

The three traditional kingdoms of multicellular eukaryotic life, plants, animals and fungi, account for most of the visible biosphere and for the majority of catalogued species on Earth. The single-celled eukaryotes have been assembled for convenience into the protists, consisting of many diverse lineages that evolved before the appearance of plants, animals, and fungi on our planet. Currently, ca. 2 million species are catalogued, most of which are metazoa, streptophytes and fungi. Protists account for less than 10% of these described species. However, considering the relative proportion of genetic protistan signatures from environmental samples and comparing these to the number of described multicellular organisms, it becomes evident that the vast majority of eukaryotic life seems to be unicellular (Pawlowski et al. 2012). As pointed out in the literature, much biological research depends upon species diagnoses. However, taxonomic expertise is collapsing (Hebert et al. 2003) and the number of undescribed protists is still in the hundreds of thousands (Pawlowski et al. 2012, Fig. 1). For technical and biological reasons, protists are usually difficult to identify for all but a few experts. The discovery and description of new species is severely hampered and many protistan lineages contain only a relatively small number of formally described species. Because barcoding has the potential to reduce ambiguities due to morphological identifications and can unmask morphologically similar species, the Consortium for the Barcode of Life (CBOL: http://www.barcodeoflife.org/) has initiated the Protistan Working Group (ProWG). The ultimate objective of the CBOL ProWG is to establish universal criteria for barcode-based species identification in protists.

In order to (i) overcome the taxonomic impediment and facilitate the identification of living species; (ii) to survey local and global diversity; (ii) to put these data into an biogeographic and palaeoecological context and to (iv) facilitate the discovery of as yet undetected species, we developed a strategy for the barcoding of ciliated protozoa. Ciliates are major components of soil and aquatic microbial communities. As most important consumers of algal biomass in lakes, as a transfer of energy and organic matter from the microbial food web to higher trophic levels and as a regulating factor of bacterial abundances and activities, ciliates play crucial roles in microbial food webs. They are capable to react within short periods of time to environmental changes making them excellent indicators of environmental health and changes and excellent model organisms.

Ciliates have the advantage of a relatively solid morphospecies concept as a basis for DNA barcoding, even though cryptic species are also known. We have tested two hypervariable gene regions, the V4 region of the small subunit (SSU) ribosomal DNA (SSU rDNA) as well as the D1-D2 region of the large subunit (LSU) rDNA as potential DNA barcode markers. We analyzed 14 species (52 different strains) of the *Paramecium aurelia*-complex, as well as ten other *Paramecium* species (14 strains in total) and found a large overlap of genetic distances in the V4 region between intra- and interspecies, suggesting that this gene fragment is unsuitable for DNA barcoding. By contrast, for the D1-D2 region, the variation within species is generally lower than divergence among (sibling) species, thus showing the characteristic “barcode gap” (Fig. 2). Because the preservation and deposition of a voucher-specimen is a prerequisite in DNA barcoding and PCR-amplification of a barcode gene relies on “destructive” sampling, we also developed a strategy to accomplish this goal (Fig. 3). The primers that were used for the PCR amplification of the D1-D2 region were also successfully tested in the laboratory for five

other classes within the phylum Ciliophora. Computer-based in silico tests with available data base sequences showed that the specific primers also target the remaining five ciliate classes.

Thus, the strategy for barcoding ciliates, an ecologically important protistan taxon group, meets all criteria defined by CBOL for a successful barcoding. Having these tools in hand, the time has come to put these tools to work in order to overcome the taxonomic impediment and to contribute to answering basic questions in ecology.

Future objectives and challenges require the test of this proposed strategy to a broad range of taxa from a variety of environmental systems. This requires the hand-in-hand collaboration between taxonomic experts to first morphologically identify species and molecular biologists to obtain and evaluate their corresponding D1-D2 barcode. Standardization of species identification and genetic identification systems will be a main task to accomplish in this collaborative effort. Furthermore, the establishment of standard barcoding protocols for uncultivated and unculturable taxa is imperative. The ProWG will work towards these goals (Pawlowski et al. 2012). In order to accomplish these goals, ProWG depends on funding sources. Including microbial eukaryotes in current and future national and international barcoding initiatives may be the biggest challenge the ProWG is facing yet.

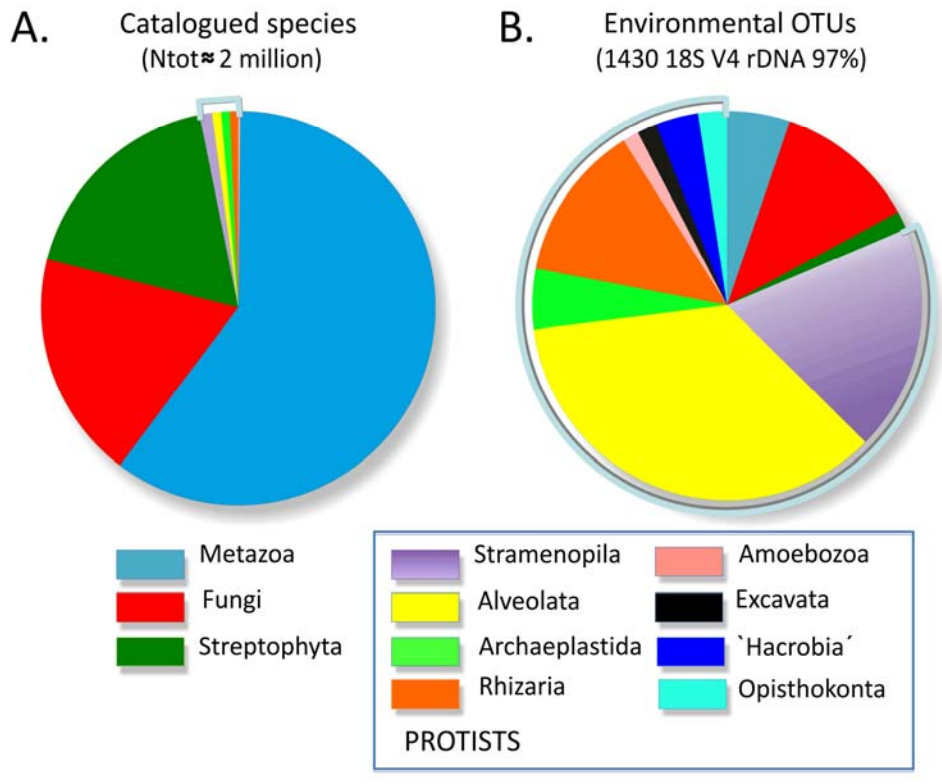

Figure 1. Morphological versus genetic views of total eukaryotic diversity. (A) Relative numbers of described species per eukaryotic supergroup. (B) Relative number of V4 18S rDNA Operational Taxonomic Units (97%) per eukaryotic supergroup, based on 59 rDNA clone library surveys of marine, fresh-water, and terrestrial total eukaryotic biodiversity. From Pawlowski et al. 2012.

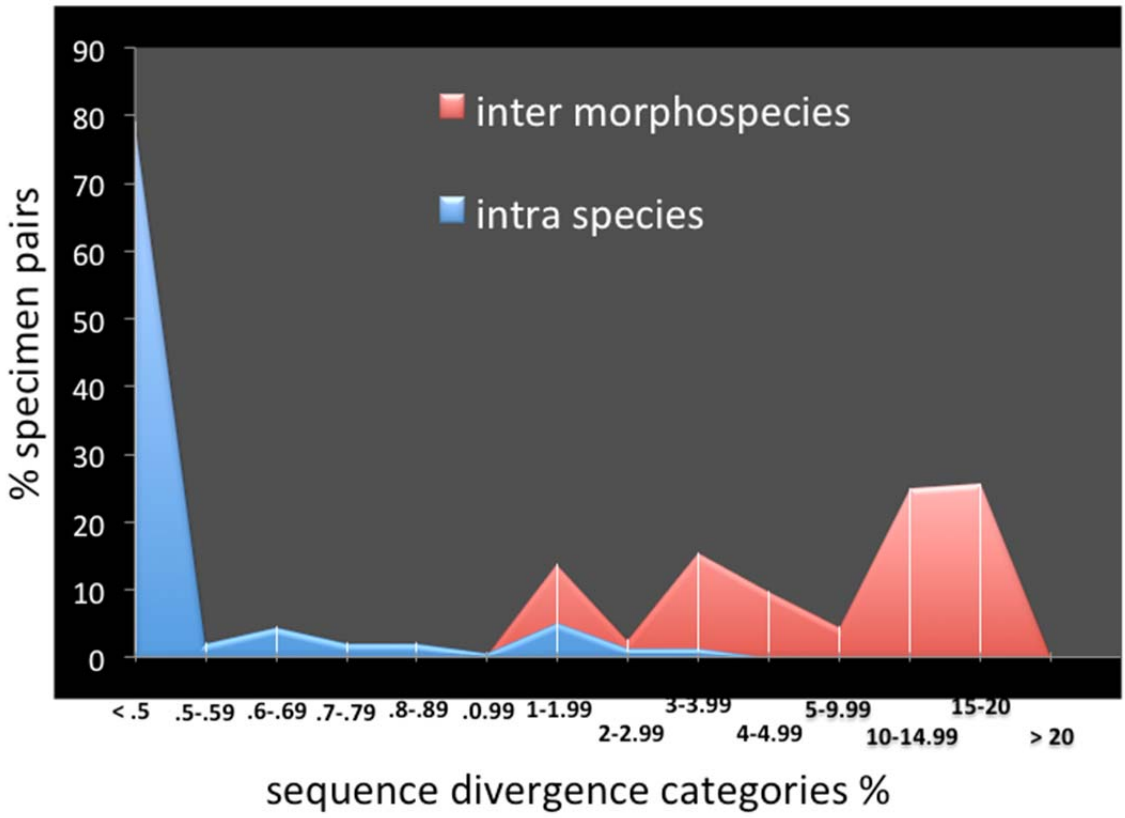

Figure 2. Divergence in D1-D2 SSU rDNA gene fragments (amplified from different Paramecium strains and species) in pairwise sequence comparisons within species (intra species) and between different species (inter morphospecies). Number of comparisons are 133 and 877, respectively. Even though a picture book ‚barcoding gap' is missing, it becomes evident that more than 90% of intra-species comparisons show less than 1% sequence divergence, whereas less than 1% of the inter-species comparisons shows a sequence divergence that falls below 1%. Thus, the D1-D2 region of the SSU rDNA shows high potential for relatively accurate discrimination of the ciliate *Paramecium* on low taxonomic level.

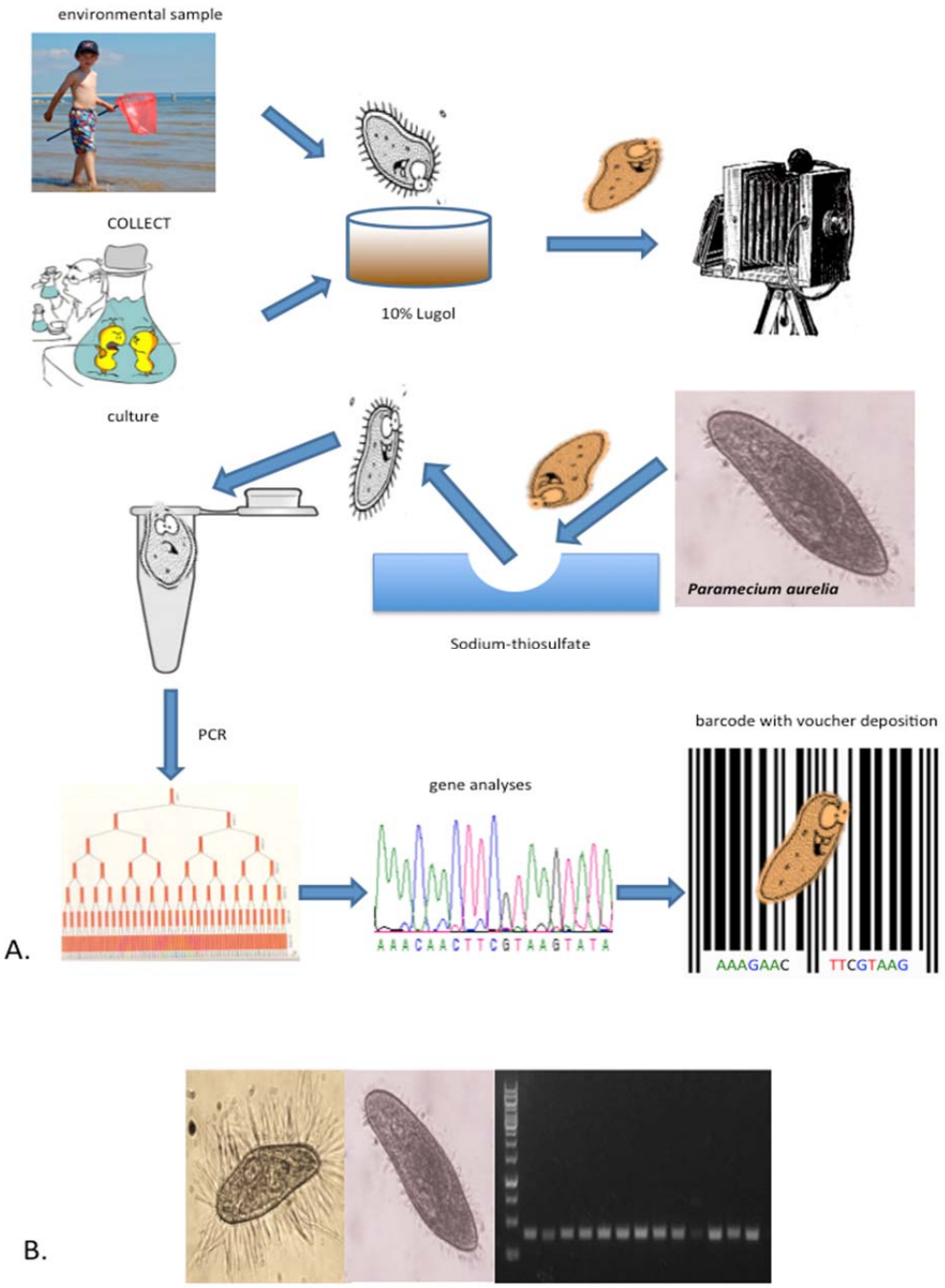

Figure 3. Strategy for voucher/specimen deposition in ciliate barcoding. (A) Cells from environmental samples or cultures are stained with acidic Lugol's solution and pictured for morphotype diagnosis. After a single cell is destained with sodium thiosulfate, this cell is subjected to gene amplification using Polymerase Chain Reaction (PCR). Subsequently, genes are analysed and all data (genes and morphological information with picture) are deposited as a voucher. (B) Lugol-stained images of two different *Paramecium* species and agarose-gel of the successful PCR-amplification of their D1-D2 SSU rDNA fragments. Cartoon culture image from Biotoon.com.

References

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003). Biological identifications through DNA barcodes. *Proc Roy Soc Lond Ser B*, **270**, 313-321.

Pawlowski J, Audic S, Adl S, Bass D, et al. (2012). CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biol*, **10** (11), e1001419. doi:10.1371/journal.pbio.1001419

Integrative taxonomy: perspectives for simultaneously improving quality and speed of a global species inventory

– Miguel Vences

University Braunschweig, Institute of Zoology, Braunschweig, Germany

Introduction

The genomic revolution has led to an enormous increase in our understanding of the evolutionary relationships of organisms. Much less attention has been paid to further develop the field of alpha taxonomy although delimiting and describing species is fundamental to define the units at which many evolutionary and ecological processes operate. I here hold the opinion that (a) the last ten years have seen various important advances that hold the potential to considerably improve taxonomic theory and practice, (b) a complete inventory of species diversity on Earth has become a feasible endeavor of high importance for science and society, and (c) that such an inventory should be done in the well-established framework of Linnean taxonomy although some of the historical work protocols of this discipline require change.

Methodological and conceptual progress in taxonomy

Despite the comparatively low research intensity of furthering the field of taxonomy, conceptual and methodological progress is apparent in four main fields:

(1) *DNA barcoding rampantly accelerated species identification and initial discovery of candidate species.* Analysis of DNA sequences provides a very efficient tool to identify organisms to species, extending to all life-history stages (eggs, larvae, juvenile, adults) and divergences in DNA sequences can also serve to flag individual organisms as candidates for representing new, undescribed species. Detailed taxonomic study can then be targeted to these candidate species (Vieites et al. 2009). The great potential of such molecular identification and discovery of species explains the enormous success of the DNA barcoding initiative (Hebert et al. 2003) which in animals uses a fragment of the mitochondrial COI gene, with about 2 million specimens and 170,000 species sequenced so far (http://www.barcodinglife.com/; accessed January 2013). The acceleration of species discovery by DNA barcoding is obvious from numerous recent studies (e.g., Fontaneto et al. 2008; Vieites et al. 2009; Fonseca et al. 2010; Poulin et al. 2011).

(2) *The General Lineage Concept of species (GLC) introduced a consensual framework for taxonomy.* During many years, discussions about species concepts have been intense, and a large number of competing concepts have been proposed. Mayden (1997) and de Queiroz (1998) noted that despite this dispute, there is fundamental agreement among all species concepts that species are evolutionary lineages indepedent from other such lineages. Under this general lineage or evolutionary concept of species, the former competing species concepts become species criteria, to be used to delimit species in practice (de Queiroz 2007). An important intellectual advance has thus been achieved, realizing that all species are independent evolutionary lineages, although not all independent lineages are species under every species criterion. Considerable disagreement might still exist to decide on the species status on a lineage, but the discussion has moved to a more practical level.

(3) *Integrative Taxonomy operates under the GLC and uses a plurality of species criteria to delimit species.* Directly derived from the understanding of species as evolutionary lineages is the understanding that not only different sets of characters (e.g., molecular and morphological) but also different kinds of arguments (e.g., population genetics, phylogeny, qualitative or quantitatve analysis of morphology) can be used to delimit species under the different species criteria available. Integration of all available evidence is therefore a powerful method to elaborate taxonomic hypotheses of high quality and reliability. Such Integrative Taxonomy (Dayrat 2005; Padial et al. 2010) means relying on the kind of evidence most appropriate to the specific group of organisms, and results in species hypotheses of increasing reliability if supported by concordance of various kinds of evidence. This does not differ from well-done taxonomy as applied historically but conceptually underpins such best practice, i.e., by defining the goal of reliably delimiting independent evolutionary lineages.

(4) *Software implementations of Integrative Taxonomy assess genealogical concordance in multilocus data sets.* One of the species criteria used to identify independent lineages, the genealogical species criterion (Avise & Ball 1990) is particularly prone to provide a fast and automated yet highly reliable means of species delimitation in the genomics age. Concordant strong differentiation in different unlinked loci provides a strong argument to assess genetic isolation and can be easily assessed by algorithms, for instance in Bayesian Assignment Tests (e.g., Leaché & Fujita 2010; Weisrock et al. 2010). Refinement of such software pipelines is promising and should include defining molecular markers of an appropriate level of variability and including other data sets (such as morphology and geography).

Prospective for a complete inventory of life

The advances summarized above have greatly accelerated species discovery and allow for an improved integrative species delimitation which however is not yet routinely used in the bulk of taxonomic studies. Nevertheless this does not yet translate into an acceleration of species descriptions. The main remaining bottleneck in taxonomy is the slow rate of taxonomic descriptions of species (ca. 20,000 per year, compared to an estimated 10 million new species of animals).

Is it useful to accelerate species descriptions and work towards a global inventory of life? Along with others (Wheeler et al. 2012) I posit that a complete inventory of global biodiversity is at the same time a fascinating scientific vision and a pressing need in the light of massive global change. It would provide the basis for numerous major questions of evolutionary biology and macroecology, and make an unprecedented variety of organisms accessible for science and conservation.

Although ambitious, such an endeavor is feasible in the light of genomic and bioinformatic advances. Proposals of a fully molecular taxonomy (Tautz et al. 2003) in which a DNA sequence becomes the primary identifier are appealing, but it will be preferable to stick to the use of the traditional Linnean taxonomic scheme which ensures historical continuity and satisfy the human preference for meaningful names rather than anonymous sequence identifiers.

In fact, Linnean nomenclatural codes are flexible enough to accommodate novel approaches. It is rather the practical work procedures of taxonomists that require reforms, and in some hyperdiverse organismal groups such as nematodes or tiny arthropods, a full "rebooting" of these protocols might be necessary. In such groups, turbo-taxonomic (Areekul-Butcher & Quicke 2012) species descriptions

can be based on automated imaging and multilocus species delimitation. Detailed anatomical study and assignment of synonyms can be partly decoupled from this procedure without violating any provisions of the nomenclature codes, and without compromising quality and value of the results.

I emphasize that the proposed approaches would be far from "quick and dirty". First, they do not apply to all groups of organisms but should be restricted to those extremely species-rich groups where a species inventory is unconceivable using traditional protocols. Second, the overall quality of the results would not be lower than in current taxonomic practice but would rather represent a shift in paradigms: (a) Detailed knowledge of anatomy of single species decreases but overall knowledge of morphological variation and morphological evolution in an entire clade increases; (b) morphological diagnosibility of species decreases but the reliability of species delimitation increases; (c) morphology-based surveys become more difficult but DNA-based surveys of species diversity, using bulk sampling or environmental DNA (eDNA) becomes feasible and will provide a realistic and fascinating picture of community dynamics, global distribution patterns and ecology of currently completely neglected clades of organisms.

Most importantly, we are living in an era of global change and facing a sixth mass extinction on Earth. In many respectes, we are running out of time. Inventorying, in a realistic timeframe of several decades, all species on Earth and obtaining realistic data on the ecology and distribution also of neglected hyperdiverse groups provides an essential baseline for political and societal decision-making. Not only will we know the ranges and requirements of all species on the globe and thereby assess spatial and taxonomic conservation priorities. By basing species delimitation increasingly on genetic and genomic data we will also provide the baseline for very efficient eDNA monitoring of shifting ranges and local adaptations with global warming, and for a true understanding of the importance of single species and species diversity for ecosystem functions and services.

References

Areekul-Butcher B, Quicke DLJ (2012). A turbo-taxonomic study of Thai *Aleiodes* (*Aleiodes*) and *Aleiodes* (*Arcaleiodes*) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. *Zootaxa*, **3457**, 1-232.

Avise JC, Ball RM (1990). Principles of genealogical concordance in species concepts and biological taxonomy. In: Futuyma D, Antonovics J (eds.). Surveys in Evolutionary Biology Volume 7. Oxford University Press, 45-67.

Dayrat B (2005). Toward integrative taxonomy. *Biol J Linn Soc*, **85**, 407-415.

de Queiroz K (1998). The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In: Howard DJ, Berlocher SH (eds.). Endless Forms: Species and Speciation. Oxford University Press, New York, 57-75.

de Queiroz K. 2007. Species concepts and species delimitation. *Syst Biol*, **56**, 879-886.

Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lambshead PJ, Thomas WK, Creer S (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Comm*, **1**, e98.

Fontaneto D, Boschetti C, Ricci C (2008). Cryptic diversification in ancient asexuals: evidence from the bdelloid rotifer *Philodina flaviceps. J Evol Biol*, **21**, 580-587.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003). Biological identifications through DNA barcodes. *Proc R Soc London B*, **270**, 313-321.

Leaché AD, Fujita MK (2010). Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proc R Soc B*, **278**, 490-492.

Mayden RL (1997). A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge M, Darwah HA, Wilson MR (eds.). Species: The Units of Biodiversity. London, Chapman and Hall, 381-424.

Padial JM, Miralles A, De la Riva I, Vences M (2010). The integrative future of taxonomy. *Front Zool*, **7**, 16.

Poulin R (2011). Uneven distribution of cryptic diversity among higher taxa of parasitic worms. *Biol Lett*, **7**, 241-244.

Tautz D, Arctander P, Minelli A, et al. (2003). A plea for DNA taxonomy. *Trends Ecol Evol*, **18**, 70-74.

Vieites DR, Wollenberg KC, Andreone F, Köhler J, Glaw F, Vences M (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proc Natl Acad Sci USA*, **106**, 8267-8272.

Weisrock DW, Rasoloarison RM, Fiorentino I, Ralison JM, Goodman SM, Kappeler PM Yoder AD (2010). Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PLoS ONE*, **5**, e9883.

Wheeler QD, Knapp S, Stevenson DW, Stevenson J, Blum SD, Boom BM, Borisy GG, Buizer JL, De Carvalho MR, Cibrian A, Donoghue MJ, Doyle V, Gerson EM, Graham CH, Graves P, Graves SJ, Guralnick RP, Hamilton AL, Hanken J, Law W, Lipscomb DL, Lovejoy TE, Miller H, Miller JS, Naeem S, Novacek MJ, Page LM, Platnick NI, Porter-Morgan H, Raven PH, Solis MA, Valdecasas AG, Van Der Leeuw S, Vasco A, Vermeulen N, Vogel J, Walls RL, Wilson EO, Woolley JB (2012). Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity. *Systematics and Biodiversity,* **10,** 1-20.

The German Barcode of Life Project and the contribution of DNA barcoding to modernize taxonomy

– J. Wolfgang Wägele

Zoological Research Museum Alexander Koenig – Leibniz-Institute for Animal Biodiversity, Bonn, Germany

The difficulty to identify organisms correctly in an efficient and quick way is an enormous impediment[1] for any research or application that requires information on rare or unknown species or on the composition of complex biological communities (Hebert *et al.* 2003; Janzen *et al.* 2009; Wägele *et al.* 2011). This difficulty is caused by

- keys for the identification of species that can only be used by trained experts
- the scarcity of such experts and their narrow taxonomic specialization which makes it necessary to involve a larger number of experts whenever environmental samples have to be studied
- the slowness of the traditional identification process

This taxonomic impediment is a serious handicap for basic ecological research, for nature conservation, veterinary medicine, pest management, fisheries, pharmaceutics, forensics etc. A consequence of this situation is that

- the quality of biological communities is described with only a few "indicator species" which never are indicators for the composition of all elements of a habitat
- frequently physiological or biochemical studies are based on misnamed species or unidentified material
- ecologists got used to the fact that they do not get the data they need, wherefore they focus on flux and cycle studies, on biomass measurements etc. which in future are useless as baseline to understand changes of species diversity (examples: (Bluhm *et al.* 2005; Glenday 2006; Kröncke *et al.* 2003; Lavelle & Pashanasi 1989; Moser *et al.* 2008)

Currently it is not possible to answer pressing questions that can strongly influence the quality of life of future generations:

- How many species are getting extinct every year, which factors are causing the fast deterioration of species diversity in our planet's biosphere?
- How is climate change and intensification of land use changing the composition of animal, plant and microbial communities?
- Which species can survive in national parks and other protected habitats?
- Are measures for nature conservation like the establishment of habitat corridors effective?

[1] See the „Darwin Declaration": http://www.cbd.int/doc/meetings/cop/cop-04/information/cop-04-inf-28-en.pdf

To be able to confine the anthropogenic damage to our planet's most important renewable resource, the German parliament has passed the National Biodiversity Strategy[2]. One of the most important demands of this strategy is the monitoring of biodiversity. Though the urgency for a national monitoring program is easily explained and comprehensible, there are currently no methods that allow the compilation of the required data, and there is no national research program to develop workflows and standards.

New methods and workflows are also needed for the IPBES process. "The Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) is a mechanism proposed to further strengthen the science-policy interface on biodiversity and ecosystem services, and add to the contribution of existing processes that aim at ensuring that decisions are made on the basis of the best available scientific information on conservation and sustainable use of biodiversity and ecosystem services" (text from the IUCN about IPBES[3]). The "best available scientific information" is still very poor: biologists do not possess tools comparable to meteorological stations to measure environmental changes.

GBOL is the first larger project funded by the German Federal Ministry for Education and Research that contributes to the development of new technologies for biodiversity monitoring. Our vision is that in future it should be possible to identify any organism occurring in Germany in short time and for little money. GBOL is a network of natural history museums and university institutes who share the workload and specialize in barcoding of different taxa. The goal is to collect all species found in Germany and to feed a central GBOL database with data on localities, genetic markers and (in future) habitat and niche characteristics. This database is the prerequisite for an automatized identification of species in biological samples. GBOL also tests workflows for practical studies, such as monitoring of freshwater benthos, identification of parasites, or analysis of the composition of the fauna in different soil types.

To complete this mission, GBOL will need support for several years. The data compiled with this project will be useful forever, because from the perspective of a human life span the documented genetic markers will not change significantly in several hundred years, and should evolutionary changes occur, these will be easily detected. Therefore, the investments are sustainable and seminal.

References

Bluhm BA, MacDonald IR, Debenham C (2005). Macro- and megabenthic communities in the high Arctic Canada Basin: initial findings. *Polar Biology,* **28**, 218-231.

Glenday J (2006). Carbon storage and emissions offset potential in an East African tropical rainforest. *Forest Ecology And Management,* **235**, 72-83.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B Biological Sciences,* **270**, 313-321.

Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J-M, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID, Harvey DJ, Hausmann A, Kitching IJ, Lafontaine D, Landry J-F, Lemaire C, Miller JY, Miller JS,

[2] http://www.bmu.de/naturschutz_biologische_vielfalt/downloads/publ/40333.php

[3] http://iucn.org/about/work/programmes/global_policy/gpu_what_we_do/ipbes/about_ipbes/

Miller L, Ratnasingham S, Miller SE, Montero J, Munroe E, Rab Green S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ (2009). Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources,* **9** (Suppl.1), 1-26.

Kröncke I, Türkay M, Fiege D (2003). Macrofauna communities in the eastern mediterranean deep sea. *Marine Ecology,* **24**, 193-216.

Lavelle P, Pashanasi B (1989). Soil macrofauna and land management in Peruvian Amazonia (Yurimaguas, Loreto). *Pedobiologia,* **33**, 283-291.

Moser G, Röderstein M, Soethe N, Hertel D, Leuschner C (2008). Altitudinal changes in stand structure and biomass allocation of tropical mountain forests in relation to microclimate and soil chemistry. In: Beck E, Bendix J, Kottke I, Makeschin F, Mosandl R (Eds.). Gradients in a tropical mountain ecosystem of Ecuador. Berlin, Springer, 229-242.

Wägele H, Klussmann-Kolb A, Kuhlmann M, Haszprunar G, Lindberg DR, Koch A, Wägele JW (2011). The taxonomist - an endangered race. A practical proposal for its survival. *Frontiers in Zoology,* **8** (25), 1-7.

"The Age of Taxonomy"

– Quentin Wheeler
International Institute for Species Exploration, Arizona State University, Tempe, USA

Technological developments like the Hubbell telescope and radiometric sensors have added to and expanded the tool kit available to astronomy. In the same way, the impact of genomics is to augment and expand the capabilities of taxonomy, not replace existing ones. The fundamental questions asked by taxonomists for hundreds of years remain as interesting and relevant as ever. The only difference is having additional tools to pursue answers.

Applications of taxonomic knowledge allow us to identify species and produce cladograms (phylogenetic "trees"), but these are services based on by-products of taxonomic research and not the driving intellectual force behind the discipline that aims instead to explore and understand the origin and evolutionary history of species, characters, and clades.

Stephen Jay Gould once said that "*H. sapiens* is but a tiny, late-arising twig on life's enormously arborescent bush — a small bud that would almost surely not appear a second time if we could replant the bush from seed and let it grow again." Evolution is repleat with contingencies — accidents of history — and does not behave with lawlike repeatability. In order to understand ourselves, where we came from, and how we fit in the natural world, we must study and reconstruct in detail the history of species and their evolutionary novelties. This is a fascinating and complex narrative that cannot be modeled or generalized, but rather a story that must be told in chronological order and in singular details. Telling this story is uniquely the task of taxonomy.

Were we able to use spectroscopy and isotope signatures we could conceivably recognize each Da Vinci painting and perhaps even date them and the geographic sources of their paints and pigments. While this might permit us to identify them and see them in their chronological order, it would not begin to tell us the things that make great paintings worthy of study to begin with. As Richard Dawkins has observed "The essence of life is statistical improbability on a colossal scale," meaning that the most fascinating things about evolutionary history are the unlikely complex characters of anatomy, behavior, development, and so forth. While genomics data by itself might permit accurate identifications and construction of trees, it is the thoughtful analysis of characters, homology, and synapomorphy that inform us of much of what we hope and need to learn about evolutionary history.

Thus, Kierkegaard's description of the history of human events applies equally well to evolutionary history: "Life can only be understood backwards; but it must be lived forwards." This is why geneticists who study populations (species in the making) tell us much about processes of speciation but are powerless to predict which isolated populations will ultimately become full blown species. It is the focus of taxonomy to understand character transformations, speciation and cladogenesis by looking backward and analyzing all sources of evidence relevant to understanding evolutionary history including fossils, anatomy, DNA, and ontogeny.

E.O. Wilson in his book *Consilience* noted that "Historians of science often observe that asking the right question is more important than producing the right answer. The right answer to a trivial question is also trivial, but the right question, even when insoluble in exact form, is a guide to major dis-

covery." Taxonomy continues to ask the right questions that will lead us to major discoveries about phylogeny.

Astronomers and cosmologists have done a remarkable job in recent decades improving our understanding of the origin and history of the Universe, as have anthropologists in filling gaps in the six-million year emergence of Modern Man. From the perspective of evolutionary history, however, the "sweet spot" is the 3.8 billion years between the appearance of the first living organism and the spectacular biodiversity that surrounds us today. Exploring this biological "cosmos" is the job of taxonomy.

Discovering, describing, classifying and naming species is as relevant as ever. Descriptions not only make biologists aware of the evolutionary novelties of species and clades, they open these adaptive innovations to human exploitation. Biomimicry will play an increasingly important role as we confront increasing numbers of environmental challenges. Natural selection has worked ceaselessly for nearly four billion years conducting countless trial and error experiments to find sustainable solutions to the same survival issues humans face today. Failure to describe species is failure to open this vast library of possibilities to society.

Because of the biodiversity crisis the tasks of taxonomy have become urgent. Many species that we do not preserve in natural history museums and describe in detail will simply be lost to science. Few will leave a fossil record behind, yet knowledge of their evolutionary novelties are critically important pieces of the story of life on earth. Fortunately, advances in cyber-infrastructure mean that the inefficiencies of the past need no longer be tolerated. Modernizing museum collections-based aspects of so-called descriptive taxonomy, along with utilization of genomics, means that we can work more rapidly than ever before. The report from a workshop held in New York called "Sustain What? Mission to Explore and Conserve Biodiversity" offers several strategies by which we could immediately accelerate the rate of species discovery and description by an order of magnitude. This would bring the rate of discovery to about 200,000 species per year which would make it possible to describe the estimated ten million species unknown to science in no more than fifty years, utterly transforming the evolutionary and environmental sciences.

Reference

Wheeler QD, Knapp S, Stevenson DW, Stevenson J, Blum SD, Boom BM, Borisy GG, Buizer JL, De Carvalho MR, Cibrian A, Donoghue MJ, Doyle V, Gerson EM, Graham CH, Graves P, Graves SJ, Guralnick RP, Hamilton AL, Hanken J, Law W, Lipscomb DL, Lovejoy TE, Miller H, Miller JS, Naeem S, Novacek MJ, Page LM, Platnick NI, Porter-Morgan H, Raven PH, Solis MA, Valdecasas AG, Van Der Leeuw S, Vasco A, Vermeulen N, Vogel J, Walls RL, Wilson EO, Woolley JB (2012). Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity. *Systematics and Biodiversity,* **10,** 1-20.

Deutsche Akademie der Naturforscher Leopoldina e.V.
– Nationale Akademie der Wissenschaften –

Jägerberg 1
06108 Halle (Saale)
Tel.: (0345) 472 39-600
Fax: (0345) 472 39-919
E-Mail: leopoldina@leopoldina.org

Berliner Büro:
Reinhardtstraße 14
10117 Berlin

Die Leopoldina ist eine der ältesten Wissenschaftsakademien der Welt. 1652 gegründet, ist sie der freien Wissenschaft zum Wohle der Menschen und der Gestaltung der Zukunft verpflichtet. Mit ihren rund 1500 Mitgliedern versammelt die Leopoldina hervorragende Wissenschaftlerinnen und Wissenschaftler aus Deutschland, Österreich, der Schweiz und zahlreichen weiteren Ländern.

Als Nationale Akademie Deutschlands vertritt die Leopoldina seit 2008 die deutsche Wissenschaft in internationalen Gremien und nimmt zu wissenschaftlichen Grundlagen politischer und gesellschaftlicher Fragen unabhängig Stellung. In interdisziplinären Expertengruppen erarbeitet die Leopoldina, auch gemeinsam mit anderen deutschen, europäischen und internationalen Akademien, öffentliche Stellungnahmen zu aktuellen Themen. Die Leopoldina fördert die wissenschaftliche und öffentliche Diskussion, sie unterstützt wissenschaftlichen Nachwuchs, verleiht Auszeichnungen, führt Forschungsprojekte durch und setzt sich für die Wahrung der Menschenrechte verfolgter Wissenschaftler ein.

www.leopoldina.org