Czech Science Foundation - Part C
Project Description
Applicant: **Doc. Pavel Škaloud, Ph.D.**
Name of the Project: **Generating the species – towards a better understanding of speciation mechanisms in eukaryotic microorganisms**

1. Background

More than 150 years since Darwin (1859) published his ideas on the origin of species, evolutionary biologists try to untangle processes which influence the speciation. Speciation is the creative engine for generating species richness; therefore, understanding the general patterns and processes of speciation is fundamental to explaining the diversity of life (Mayr 1963). The expansion of molecular phylogenetics over the past decades has opened up a powerful new approach to study the speciation mechanisms. Phylogenetic trees, particularly those including all the living species in a higher taxonomic group, provide an indirect record of the speciation events that have led to present-day species (Hennig 1966). Together with information on the geographical and ecological attributes of species, they can provide information on the causes of speciation within the group of interest (Barraclough & Nee 2001).

Though eukaryotic microorganisms (protists) are extremely numerous, diverse and essential in global ecosystem functioning, they are largely understudied by evolutionary biologists compared with multicellular organisms (Gerstein & Moore 2011). In part owing to their small sizes and difficulty in culturing, our knowledge of their diversity and evolutionary processes is considerably limited. In particular, very little is known about speciation mechanisms generating the protist diversity. Yeasts are the most investigated organisms in this respect, which is mostly caused by the availability of their population genomic data (see Louis 2011 for review). However, similar investigations on non-fungal protists are extremely scarce. Nevertheless, some of the molecular studies provided examples of limited dispersal and possibilities for both allopatric (Evans et al. 2009, Degerlund et al. 2012) and sympatric speciation (Amato et al. 2007, Poulíčková et al. 2014). Allopatric speciation has been corroborated by a study presenting an experimental evidence of limited dispersal capacities in diatoms (Souffreau et al. 2010). Fossil records point to the sympatric speciation of planktonic protists (Lazarus 1983, Benton & Pearson 2001).

To understand the mechanisms that have generated and distributed the species diversity, it is necessary to investigate evolutionary processes occurring at species and population level, using sufficiently discriminatory molecular markers. Differentiation between populations is a necessary step in speciation, during which reproductive isolation barriers arise. Recent population studies of free-living aquatic protists mainly focused on diatoms and dinoflagellates. These studies showed spatial (Casteleyn et al. 2010, Evans et al. 2009) as well as temporal differentiation of populations, which may lead to ecological constraints to gene flow (D'Alelio et al. 2009, Rynearson et al. 2006). Population differentiation has also been shown in several other phototrophic microorganisms, such as prymnesiophytes (Iglesias-Rodríguez et al. 2006), raphidophytes (Kooistra et al. 2001) and dictyochophytes (Riisberg & Edvardsen 2008). The mechanisms of genetic differentiation are still not well understood. However, it has been shown that spatial variation in protist communities can be predominantly explained by specialization to habitats (Logares et al. 2007, 2008, Gächter & Weisse 2006, Lowe et al. 2005) and historical factors (Vyverman et al. 2007). **Therefore, it is very likely that ecological differentiation and dispersal capacities represent the driving forces behind the speciation mechanisms of protists.** However, empirical studies testing these hypotheses using robust datasets are virtually missing (De Gelas & De Meester 2005).

2. The scientific aims of the project

We propose an innovative, complex research program that has a very high probability of substantially advancing our **understanding of the speciation mechanisms in eukaryotic microorganisms**. We hope that our findings will open entirely new ways for investigation and conservation of the planktonic protist communities. The ground-breaking nature of the project lies in the complexity of approaches utilized (broad field sampling, population genetics, ecophysiological experiments, niche modelling, dispersal capacity models) and in application of the innovative, cost-effective approach of reduced-representation libraries enabling us to molecularly characterize the population structure based on the genome-wide genotyping of hundreds of isolated organisms.

We will use the planktonic chrysophyte genus *Synura* as a model organism, specifically the species belonging to the *S. petersenii* group (Škaloud et al. 2014). This group of species represents one of the most widely distributed and common groups of freshwater microorganisms. It is relatively easily cultivated and molecularly well characterized (Boo et al. 2010, Škaloud et al. 2012, 2014), so it represents an ideal model taxon for investigating evolutionary patterns in protists. In addition, the recent investigations indicate on-going speciation and differentiation of evolutionary young lineages (Jo et al. 2013, Škaloud et al. 2014), giving us an ideal opportunity to study speciation processes, and to **determine whether niche differentiation and limited dispersal are the main promoters of speciation**.

In the following, we will present the **three specific objectives of the project** that aim to examine i) the population structure, ii) niche differentiation, and iii) dispersal capacities of *Synura petersenii* group species. To identify traits important in speciation, data obtained in these objectives will be integrated to **bring the population genetics into an ecological, geographical and temporal context**.

Objective 1: Population structure of *Synura* species

Distribution patterns of species and populations are driven by an assemblage of factors, including abiotic (climate, geography, geology) and biotic factors (dispersal limitations, interactions among species; Araújo & Guisan 2006). Unveiling the overall distribution and population structure of species represents a fundamental and essential step towards understanding the mechanisms of protist speciation. Dispersal and distribution of microscopic eukaryotes, has become a highly controversial topic in recent years. Opposing views have asserted that these organisms are composed of many species that have limited geographical distributions on the one hand (Foissner 1999), or whose distributions are cosmopolitan on the other (Fenchel & Finlay 2004). Recent studies, however, strongly support what Foissner (1999) proposed, which allow us to finally reject a generalized ubiquitous dispersal hypothesis for microorganisms (van der Gast 2015).

Several recently published studies have reported the distribution patterns of species within the *S. petersenii* group (Kynčlová et al. 2010, Boo et al. 2010, Škaloud et al. 2012, 2014). In general, a broad range of distribution patterns have been reported, ranging from broadly cosmopolitan to extremely endemic species. This knowledge allows us to investigate the population structure, and subsequently the speciation mechanisms, on four selected species exhibiting four different distributions: *S. petersenii* (broadly cosmopolitan), *S. glabra* (cosmopolitan, thermophilic), *S. americana* (common in North America, rare in Europe), *S. hibernica* (endemic to western Ireland). Our pilot studies based on ITS rDNA data show the existence of population differentiation in all four selected species, even indicate the existence of limited gene flow among the climatic regions. We will ask the following questions:

Q_1. How are the populations structured? Are the populations geographically and temporally differentiated? How a demography of the populations varied over time?
Q_2. How do the species restricted in their distribution differ from the cosmopolitan ones in their population structure?

Objective 2: Determining realized and fundamental niches

Ecological-niche theory is central to understanding the spatial distribution of populations. It defines a realized niche as the actual space that an organism inhabits in a biological community, and a fundamental niche as a set of physiological conditions that allows individual to survive and reproduce in absence of biotic interactions (Keller & Lloyd 1994). Habitat or niche models are commonly used to link the distribution of species to variation in a series of environmental variables. The fundamental niche can be estimated by direct measurements of physiological limits (a mechanistic approach), whereas the realized niche is usually determined by relating data on occurrences with environmental data sets (a correlative approach) (Soberón & Peterson 2005).

In Objective 2, we will focus on niche differentiation of particular genotypes by both mechanistic and correlative approach, to investigate the importance of local adaptation in the divergence of incipient species. First, isolated and cultivated strains representing different genotypes will be subjected to various experimental studies estimating their ecological amplitudes with respect to

the light regime, temperature, and nutrients. It has been shown that light and temperature represent the most important abiotic factors in aquatic environments, as they present great variations on a spatial as well as on a temporal scale (Bolsenga & Vanderploeg 1992, Masclaux et al. 2009). In addition, nutrient conditions were hypothesized to represent one of the primary factors in *Synura* speciation (Škaloud et al. 2014). Second, we will collect environmental data of all sampling sites to determine the realized niche by a correlative approach. In addition to in-situ measured pH and conductivity (approximately reflecting the availability and amount of nutrients), various bioclimatic variables will be acquired, as well. We will ask the following questions:

Q_3. How do the genotypes differ in their physiological responses to major environmental factors?
Q_4. Are the particular genotypes adapted to local environmental conditions? E.g., are the temperature growth optima of Northern populations shifted towards the colder temperatures?

Objective 3: Evaluating dispersal capacities

Speciation is a complex process involving multiple contributory factors. Along with the abiotic factors, speciation processes may be influenced by various biotic factors, including competition, predation, and symbiotic interactions. In addition, the integration of biotic and abiotic factors over time (i.e., niche pre-emption) may be important in creating population patterns, but is not well-studied yet (Wiens 2011). In freshwater diatoms, the sensitivity to freezing, desiccation and abrupt heating was demonstrated to significantly influence their dispersal capacities and consequently rates of allopatric speciation (Souffreau et al. 2010). Similarly, restricted distribution patterns of several *Synura* species point to the significant role of dispersal in the speciation of these organisms (Škaloud et al. 2014).

In Objective 3, we will focus on dispersal capacity as an important biotic factor influencing the gene flow, and such the speciation processes in freshwater protists. First, we will compare the realized and fundamental niches of particular *Synura* genotypes obtained in Objective 2. Differences in these two niche models would reflect the biotic factors in determining the genotype distributions, including the dispersal capacities. To test for the net effect of dispersal limitation on the niche models, we will experimentally investigate the tolerance to desiccation and temperature stress, as well as the ability to form resistant cysts, in the *Synura* genotypes showing contrasting overlaps in niche models. We will ask the following questions:

Q_5. Do dispersal capacities represent the most important biotic factor in speciation processes? Are the genotypes with small realized niches greatly limited in their dispersal?
Q_6. What is the major factor limiting the dispersal capacity?

3. Relevance and topicality of the project

Species are the fundamental units of microbial diversity. Understanding of speciation will facilitate the formulation and testing of hypotheses in the most important questions facing biology today, including the fit of organisms to their environment and the dynamics and patterns of organismal diversity.

It is now widely accepted that to better understand the ecology of eukaryotic microorganisms and their roles in ecosystem functioning, it is essential to find a way how to apply the traditional ecological theories to microorganisms (Carbonero et al. 2014). During the last two decades, several studies emphasized the substantial differences between the macro- and microorganisms, including the contrasts in population sizes, dispersal, extinction rate, and speciation (de Meetser et al. 2002, Fenchel & Finlay 2004, 2006, Fenchel 2005). Our extensive knowledge of all these processes is essential in the general ecological theories and models, usually applied in macroorganismal research. The uniqueness, and hence our poor knowledge, of these processes in microorganisms thus prevent us apply the macroorganismal ecology models, and, to reconcile the ecological principles of traditional and microbial ecology.

However, several recently published studies demonstrated that the basic evolutionary processes of eukaryotic microorganisms may be comparable to multicellular animals and plants. First, there is growing evidence that microorganisms follow the same patterns as multicellular organisms when it comes to population structure and levels of genetic diversity (Rengefors et al. 2012). Second,

biogeographic studies point to the limited geographical distributions of many microbial species/lineages (van der Gast 2015). Third, the estimates of effective population sizes of microorganisms have been shown to be similar in magnitude to values reported for multicellular organisms (Watts et al. 2013).

The similarity of microbial and macroorganismal speciation mechanisms remain unresolved so far. Therefore, our project represents a substantial challenge to the worthwhile cause of incorporation the traditional ecological principles and theories into microbial research. Speciation has been a major focus of evolutionary biology research in recent years, with many important advances (Butlin et al. 2012). Understanding the role of general speciation mechanisms, in particular the niche differentiation and dispersal limitation, for microbial speciation processes will considerably improve our abilities to study the ecology and ecosystem functioning of eukaryotic microorganisms. The great promise is that this would help us better understand and predict changes in the natural environment, would allow improved ecological quality assessment of surface waters and conservation management, and would help explaining ecological phenomena at higher levels of biological organization.

4. Methodical and conceptual approaches

For achieving our goals specified above, we will use the following workflow, consisting of i) sampling, isolating and culturing *Synura* colonies, ii) selection of appropriate molecular markers, iii) genotyping, population genetics analyses, iv) experimental analyses, and v) proposing and testing speciation models (**Fig. 1**):

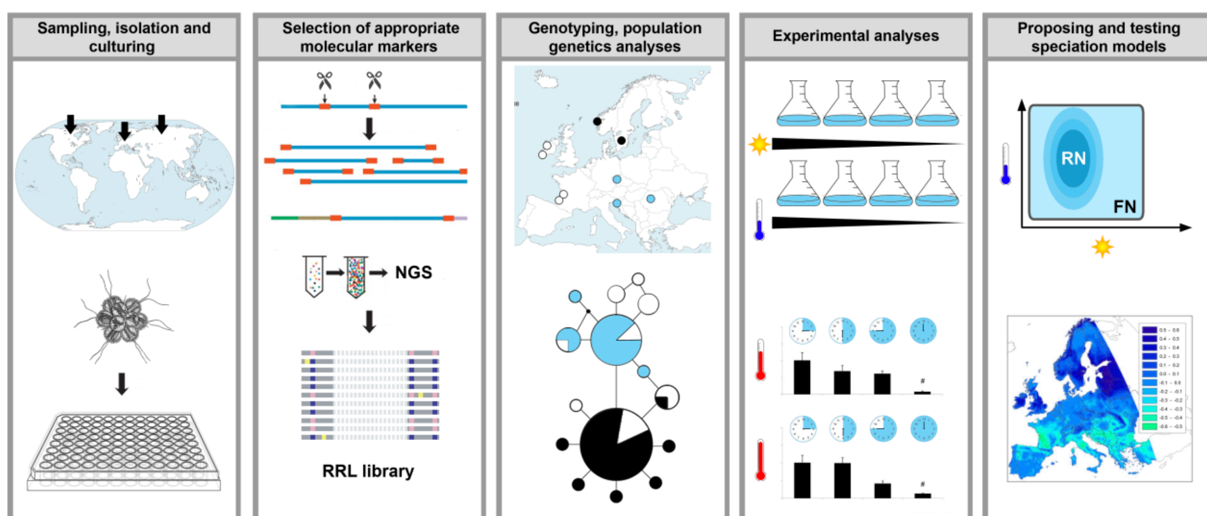

Fig. 1. A proposed integrated methodical approach for the study of genetic diversity and speciation processes in the genus *Synura*. Focusing on the microevolutionary (WP1) and population (WP2) levels is indicated.

i) ***Sampling design, isolation and culturing*** – Our research will focus on four *Synura* species, selected to represent cosmopolitan species with and without thermophilic preferences, combined with species with limited geographical distributions. In addition, the species were also selected to enhance our ability to morphologically discern them with light microscopy (the specific discriminative features include lanceolate cells of *S. petersenii*, broadly oval, tightly appressed cells of *S. glabra*, significantly elongated cells of *S. hibernica*, and almost spherical cells of *S. americana*). Field collections will be primarily concentrated in Northern temperate and boreal regions, harbouring a large degree of global *Synura* diversity, including the selected species (Kristiansen & Preisig 2007, Škaloud et al. 2013b, 2014). To cover the area of study, we will sample in Europe, North America and North Asia. In Europe, we will focus on yet genetically unstudied regions (Eastern and Western Europe; see the maps in Škaloud et al. 2014), which will result in detailed coverage of the whole Europe. The other two continents will be sampled more patchily; however, this design allows us to study the population differentiation in two geographical scales. The sampling effort will be invested in 2017 and spring 2018, as follows. In the spring of 2017, the samples will be obtained in the localities in France

(Aquitaine and Vosges), and Romania (Cluj-Napoca, near Sălicea), respectively, known for rich diversity of scaled chrysophytes (Péterfi & Momeu 2009, Němcová et al. 2012). We will perform a detailed sampling in Ireland, as well, to map the population structure of endemic species *S. hibernica*. In the autumn of 2017, we will sample in Michigan, USA and Western Ontario, Canada (Kling & Kristiansen 1983, Wawrzyniak & Andersen 1985, Wujek & Igoe 1989). The sampling in spring 2018 will be conducted in the Slovenia (Němcová 2014), North Ukraine (Matvienko 1965), Alaska, USA (Asmund 1968) and Khanty-Mansiysk (Mukhrino Field Station), Russia. The localities will be selected to cover a broad range of water body types (differing in size, depth, topography, turbidity, reciprocal remoteness, trophy, and pH). At each locality, selected environmental variables (pH, conductivity) will be measured. The strains will be isolated using the single-cell pipetting method into the 96-chambred multi-well plates filled with DY IV medium (Andersen et al. 1997), up to several hours after the field sampling. The plates will be stored and transferred in the portable cool box equipped by LED diodes. After returning to the laboratory, the strains will be cultivated at 15°C, as described in Škaloud et al. (2014). According to our previous sampling experiences, we suppose to sample ca 80 localities, and establish ca 100-120 cultures during each of the sampling events, resulting in establishing of ca 1,000 cultures in total. Our long-term experience with isolating *Synura* colonies guarantee the high feasibility of proposed sampling design. To confirm their taxonomic identity, all strains will be characterized by ITS rDNA Sanger sequencing, as described in Kynčlová et al. (2010).

ii) ***Selection of appropriate molecular markers*** – To characterize the population structure genetically, we will apply a state-of-the-art method using NGS technologies to get a high number of variable, single-copy nuclear loci. The innovative, cost-effective approach of reduced-representation libraries (RRLs; Lemmon & Lemmon 2012, Myles 2013) will be used to obtain nucleotide sequences efficiently from hundreds of orthologous nuclear loci, without the requirement of using a reference genome. Indeed, RRLs represent an ideal tool for our research, enabling us to obtain a robust, genome-wide characterization of a high number of genotypified strains. In comparison to the traditional methods of population genetics (microsatelites, AFLPs, RFLPs), RRLs are much more precise as they produce multistate data and do not require subjective scoring of marker panels. Thus, RRLs may be used for single nucleotide polymorphisms (SNPs) analysis and phylogeny based on multistate data. Comparing to the other NGS methods for genome-wide genotyping (RAD-Seq, CRoPS, MIDs), loci selection based on the levels of SNPs enables to genotype more samples at the same coverage, favouring the RRLs approach when investigating a high number of strains (Davey et al. 2011, Lemmon & Lemmon 2012). Briefly, the RRLs will be prepared by digesting genomic DNA of eight chosen cultures (two per each species) using a selected restriction enzyme. The selection of the enzyme will be based on comparing digestion performance and fragment-length distributions of digested DNA generated by five various enzymes (*Xba*I, *Eco*RI, *Eco*RV, *Bst*EII, *Not*I). After digestion with the selected enzyme, digested DNA will be sorted by size on a gel, fragments of 400-600 bp will be excised, and gel extraction will be performed. Resulting libraries will be quantified using a Qubit 2.0 fluorometer (Life Technologies), and sequenced on Illumina MiSeq platform, using the bidirectional sequencing approach. Raw data will be filtered and trimmed based on the quality of reads. After assembling reads into single-copy loci for each four cultures, sequences for each locus will be aligned. Primers amplifying ~330 bp regions will be designed for ca 200 loci selected to exhibit levels of sequence divergence useful for our population-level investigations, with the focus to avoid interacting different primer pairs. Finally, primers will be tested on a set of 24 *Synura* cultures (6 per each species), preferably isolated from a wide range of habitats and geographical regions. To reduce the sequencing costs, PCR testing will be performed using a multi-level approach as described in Lemmon & Lemmon (2002). Ca 50-80 primer sets successfully amplified their target products in all 24 cultures will be selected for subsequent population genetics analyses.

iii) ***Genotyping, population genetics analyses*** – A set of RRL primers will be used to genetically characterize each of the isolated strain. For each population, at least 10 isolated strains will be investigated. In addition to the newly isolated strains, DNA samples of 184 *S. petersenii*, 112 *S. glabra*, 55 *S. americana*, and 12 *S. hibernica* previously isolated strains will be utilized, as well. These strains were obtained during our extensive European sampling during the last six years (Škaloud et al. 2014), and the applicability of these stored DNA samples has actually been verified by successful obtaining of ITS rDNA sequences in 10 chosen strains. The RainDrop Digital PCR System will be applied to simultaneously amplify 50-80 RRL nuclear loci per each strain, using the microdroplet PCR technique (Mamanova et al. 2010). The PCR products will be purified and quantified using a Qubit 2.0 fluorimeter. The libraries will be produced using the The NEBNext® DNA Library Prep Master Mix

Set for Illumina (New England BioLabs), and multiplexed up to 96 by attaching unique barcodes. We will use the protocol and indexing oligo sequences published by Meyer & Kircher (2010). The libraries will be sequenced on a Illumina MiSeq sequencing system, using pair-end protocol due to the lowest sequencing cost per sequencing read and at the same time satisfactory data quality. Library quality control will follow standard procedures including quality assessment by Agilent 2100 Bioanalyzer (Agilent Technologies). Samples will be pooled in equimolar concentrations and library will be deliver to sequencing facility for sequencing on MiSeq platform employing pair-end 2 x 300 bp read long sequencing run. One sequencing plate – flowcell on MiSeq platform is capable of accommodating 96 samples. Considering approximate output of 15-20 million reads and therefore >150 000 reads per sample, the sequencing will result in a high coverage of RRL nuclear loci, which is required in analyses of sequentially similar genotypes in population studies. Given the expected number of 1,300 samples analyzed (incl. DNA samples stored from previous investigations) and capacity of 96 samples per plate, the number of runs will be reaching 14. Raw reads will be separated using the specific barcode, then trimmed and filtered based on quality. Aligned pair-ended reads will be assembled and mapped using the sequence of investigated loci.

Homologous RRL loci will be aligned among populations, and the resulting dataset will be analysed using maximum parsimony, maximum likelihood, and Bayesian methods, as described previously (Škaloud et al. 2014). Population structure will be explored using the program STRUCTURE (Pritchard et al. 2000). To investigate the amount nucleotide diversity, two estimators of θ will be calculated: nucleotide diversity (π) and Wattersons' θ (Hudson 1990). Population structuring will be tested using Snn statistics (i.e., nearest-neighbor statistics) (Hudson 2000) implemented in DnaSP v 5.10.01 (Librado & Rozas 2009). This statistic indicates the frequency with which nearest-neighbour sequences are found in the same group. Geographic differentiation of populations will be investigated by performing Isolation-by-distance (IBD) analyses using the IBDWS 3.23 program (Jensen et al. 2005). Furthermore, to put geographical differentiation into the time frame, we will use molecular clocks for dating of the concatenated multi-locus phylogeny reconstructed based on maximum likelihood, maximum parsimony optimality criteria or Bayesian inference in e.g. MrBayes (Ronquist & Huelsenbeck 2003) and MEGA (Tamura et al. 2013). Prior to concatenation and phylogeny, selected loci will be tested for a phylogenetic congruence using several approaches (Leigh et al. 2011). Chronograms will be reconstructed in r8s (Sanderson 2003) and BEAST (Drummond et al. 2012) using fossil and molecular calibrations (Boo et al. 2010). Geographical distribution will be plotted over the chronogram by ancestral state reconstruction of geographical distribution using e.g. Mesquite (Madison and Madison 2015) or SPREAD (Biejelec et al. 2011). A reconstruction of the demographical variation over time will be reconstructed based on coalescent theory using Bayesian skyline plot (Drummond et al. 2005) implemented in BEAST (Drummond et al. 2012).

iv) *Experimental analyses* – To investigate processes leading to species diversification, we will study the physiological differentiation of isolated clones. Assuming the ecological differentiation and dispersal capacities to be the driving forces behind the speciation mechanisms of protists, we will experimentally investigate i) the optimal growth conditions in gradients of light, temperature, and nutrients, ii) the tolerance to desiccation and temperature stress of selected strains, and iii) the ability to form resistant cysts. The growth rates will be measured in batch cultures grown under different illumination (10 - 110 $\mu mol.m^{-2}.s^{-1}$), temperature (5 - 25°C), and nutrient (original - 10-fold diluted DY IV medium) gradients. At least 5 strains from each selected genotype will be tested. Growth of algal strains will be measured by recording F_0 values at suitable intervals, using a closed FluorCam FC 800-C (Photon System Instruments) equipped by a CCD camera and four fixed LED panels. During the exponential phase, the growth curve will be used to determine the maximal growth rate. The tolerance to temperature stress will be tested on exponentially grown cultures, on three different densities. Various heating treatments will be used, differing by the stress temperature (28-38 °C) and duration of treatment (15 min-3 h). Experiments will be performed in a Labio crossed gradients unit (Czech Republic). In addition, tolerance to freezing will be tested by placing the cultures in a freezer at -20 °C for 1-4 hours (using a repeatable -1°C/min. cooling rate by applying Mr. Frosty™ freezing containers). For the desiccation treatment, the cultures will be transferred on an ultra-thin porous substrate in a 96-well filter plate, using the Phycomat system (Nowack et al. 2005). The cultures will be maintained in non-aqueous environments, under the saturated humidity conditions. Different durations of treatment (2 hours-10 days) will be used. Preliminary results revealed clear differences in the ability of various *Synura* cultures to survive under these non-aqueous conditions, with *S. petersenii*

showing a good survival in palmella stage. After all above-mentioned treatments, the cultures will be returned to standard conditions, and their viability will be tested by growth monitoring using a FluorCam fluorometer. To test the ability to form resistant cysts, the cultures grown in high temperature, a nitrogen-depleted medium, and a high cell concentration, will be regularly checked for the cysts production (Sandgren & Flanagin 1986).

v) ***Proposing and testing speciation models*** – Based on occurrence data, in-situ measured pH and conductivity values, 19 WorldClim bioclimatic variables (for more information, see http://www.worldclim.org/bioclim), and data of experimental studies, we will predict the fundamental and realized niches of particular genotypes by habitat modelling. Bioclimatic variables will be extracted using the "rgdal" and "raster" packages in R, getting mean values from all raster cells found in a radius of 5 km around each point location. For habitat modelling, we will apply a progressive Bayesian site occupancy models with imperfect detection (Gelman et al. 2003). Parallel to accurate predictions of species distributions along environmental gradients, this method allows us to estimate the probability of detection using a season as a covariate (Rota et al. 2011). The calculations will be performed using JAGS (Plummer 2003) and the R package rjags will be used to call JAGS, export results in R and calculate Bayesian occupancy models. In addition, European distribution data will be used to predict the distributions by several other modelling techniques (ANN, CTA, GAM, GBM) using BIOMOD (Thuiller et al. 2009), using the same environmental data as described above. In BIOMOD, model outputs can be translated into presence–absence maps using the optimizing thresholds. The analyses will be performed using the "biomod2" library in the R software. Based on the distribution data in Northern temperate and boreal regions, we will also create the dispersal capacity model, by simple estimating a geographical area of the minimum convex polygons containing all localities where the species has been detected. Finally, the accuracy of habitat- and dispersal-based models will be tested by comparisons with the results of our experimental analyses.

5. Time schedule

The general schedule of the project is shown in **Fig. 2**. We present the duration of the particular conceptual approaches, and key output milestones including the papers and presentations given at international phycological congresses (for detailed characteristics of publications, see the chapter 6).

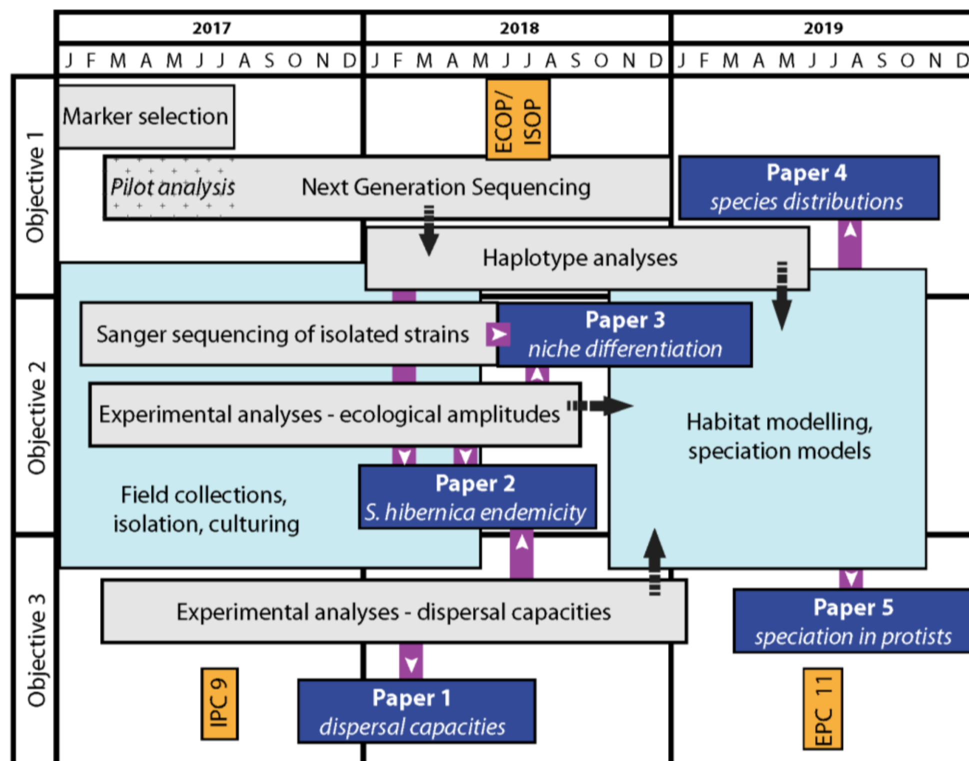

Fig. 2. Flowchart of the time schedule, including the anticipated project outputs (publication outputs in blue, presentation on congresses in orange).

6. Deliverables of the project

We plan to publish at least 5 scientific publications, as specified below. At the end of 2017, we will prepare a manuscript reporting the dispersal capacities experiments. We will try to answer the question whether the different distribution patterns of four investigated species correlate with their dispersal capacities (Fig. 2 – Paper 1: *dispersal capacities*). At the beginning of 2018, we will prepare a manuscript dealing specifically with the distribution of *S. hibernica*. We will present our results concerning the population structure, dispersal capacities, and niche differentiation of particular genotypes, and present our hypothesis clarifying the extremely restricted distribution of this species (Fig. 2 – Paper 2: *S. hibernica endemicity*). In mid-2018, after our sampling effort will be finished, we will prepare a paper focusing on niche differentiation of three remaining investigated species, applying data obtained by both mechanistic (experimental analyses) and correlative approach, to investigate the importance of local adaptation in the divergence of incipient species (Fig. 2 – Paper 3: *niche differentiation*). Finally, in 2019 we will prepare two manuscripts summarizing our results in a general context. First, we will prepare a manuscript describing the species distribution predictions, based on the habitat- and dispersal-based models (Fig. 2 – Paper 4: *species distributions*). Next, we will prepare a general paper discussing the speciation mechanisms in protists. We will focus on the geographical, temporal, and ecological differentiation of populations, describing the shifts in ecological preferences of particular haplotypes. Accordingly, we will propose, compare and test the speciation models (Fig. 2 – Paper 5: *speciation in protists*). We will select suitable, highly reputed phycological/microbiological journals for submissions of the first three above-mentioned manuscripts (e.g., Journal of Phycology, Environmental Microbiology, FEMS Microbiology Ecology). In addition, two general papers focusing on the models of species distributions (Paper 4) and speciation mechanisms in protists (Paper 5) will be published in more general, highly impacted journals (we focus on the PNAS and ISME journals). In addition, the results of the project will be presented at different phycological and protistological congresses and symposia including the the 9th Inter0national Phycological Congress (2017), the 20th ECOP-ISOP Joint Protistology Meeting (2018), and the 11th European Phycological Congress (2019).

Last but not least, the project will significantly contribute to scientific development of the young and dynamic research team, including several undergraduate and Ph.D. students. Their participation should enhance their scientific and technical knowledge in several important fields of recent biology (molecular diversity analyses, statistical data analysis) and thus boost their scientific carriers.

7. International cooperation

We have established an international cooperation with **Jørgen Kristiansen** (Copenhagen University, Denmark), a leading expert in the taxonomy of silica-scaled chrysophytes. The cooperation resulted in the publication of a review paper on the taxonomy of silica-scaled chrysophytes (Škaloud et al. 2013a). We are more than happy that Jørgen is still willing to collaborate in the subsequent studies. In the area of NGS data analysis, we will cooperate with **Edvard Glücksman** (University of Exeter, UK), an expert in next-generation-sequencing of protists. **Andreas Holzinger** (University of Innsbruck), with whom we established a fruitful cooperation a few years ago (Pichrtová et al. 2013), will be helpful in the field of eco-physiological studies.

8. Research team

The research team consists of seven researches from two phycological laboratories in the Czech Republic. The **principal investigator** (Pavel Škaloud, Charles University) gained extensive experience and recognition in algal taxonomy, in particular of green algae and chrysophytes (e.g., Škaloud et al. 2012, 2013a, b, 2014). Recently, he published several papers focusing on algal speciation mechanisms (e.g., Ryšánek et al. 2016a, b). The **co-investigator** (Petr Dvořák, Palacký University) is a leading Czech phycologist in the field of genomics of autotrophic microorganisms (Dvořák et al. 2014, 2015). Therefore, we believe that our joint experience constitute a solid base for the proposed integrated multi-approach project concentrated on chrysophycean genus *Synura*.

The research team will be headed by the **principal investigator** (Pavel Škaloud), who will be fully responsible for achieving the project objectives, project management, and overall guidance of the research team. He will also co-ordinate the molecular phylogenetic works and field collections, and will substantially participate on data analyses, presentation and publication of results. One **postdoc** (Magda Škaloudová) will co-ordinate the experimental analyses, and will participate on data analyses and publication of results. The team will be further composed of four PhD students, whose roles will

be as follows. **One PhD student** (Dora Čertnerová) will be responsible for Sanger sequencing and creating the RRL library, and will participate on experimental analyses, as well as on evaluation and publication of results. A **second PhD student** (Helena Bestová) will be responsible for statistical analyses, in particular Bayesian site occupancy, habitat- and dispersal-based models, and will participate on field collections and molecular work. A **third PhD student** (Martin Pusztai) will participate on field collections, strain isolation and culturing, data synthesis and manuscript writing. One **technician** will deal with laboratory work connected with DNA isolation, PCR, sequencing and library preparation. **Co-investigator** (Petr Dvořák) will co-ordinate the NGS work, he will be responsible for raw NGS data handling, genotyping, population genetic analyses, phylogeny, demography, and testing speciation models. His team will have **one PhD student** (Eva Jahodářová), whose responsibility will lie in cooperation on population genetics and phylogeny data analysis.

9. Facilities and equipment

The phycological laboratory at the Charles University is fully equipped to handle the isolation, cultivation, and experimental activities. The basic equipment includes Olympus and Nikon microscopes (incl. Nomarski differential contrast, phase contrast and fluorescence and microphotographic equipment Olympus U-CMAD3), the flow-box BHL 65, sterilisation autoclaves, temperated light:dark cycle controlled cultivation boxes, the spectrophotometer Spekol 1300 (Analytik Jena), several Phycomat units, the Labio crossed gradients unit, the Multi-Cultivator MC 1000 (Photon System Instruments), and the closed FluorCam FC 800-C (Photon System Instruments). The DNA laboratories at the Charles University and at the Palacký University are fully equipped with current state-of-the-art equipment to carry out the isolation of DNA, Sanger sequencing, generation of RRL libraries, DNA purification and quantification, and preparation of sequencing library for NGS. The NGS sequencing will be performed in the Institute of Experimental Botany AS CR Olomouc and the Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc. The RainDrop Digital PCR System will be available at Charles University at the end of 2016, and will be purchased from institutional funds. As extensive computational resources are needed for this project (assembly and analyses of NGS data) the large computer clusters will be used at MetaCentrum (http://metavo.metacentrum.cz/en), as well as at the Department of Botany, Palacký University (the HP Z820 workstation for bioinformatics equipped by 2x6 cores Intel Xeon, 128 GB RAM, 2x 2 TB disk space).

10. References

Amato A. et al. (2007): *Protist* 158: 193-207.
Andersen, R.A. et al. (1997): *J. Phycol.* 33 (suppl.): 1-75.
Araújo, M.G. & Guisan, A. (2006): *J. Biogeogr.* 33:1677-88.
Asmund, B. (1968): *Hydrobiologia* 31: 497-515.
Barraclough, T.G. & Nee, S. (2001): *Trends Ecol. Evol.* 16: 391-9.
Benton, M.J. & Pearson, P.N. (2001): *Trends Ecol. Evol.* 16: 405-11.
Bielec, F. et al. (2011): *Bioinformatics* 29: 2910-12.
Bolsenga, S.J. & Vanderploeg, H.A. (1992): *Hydrobiologia* 243/244: 95-104.
Boo, S.M. et al. (2010): *Mol. Ecol.* 19: 4328-38.
Butlin, R. et al. (2012): *Trends Ecol. Evol.* 27: 27-39.
Carbonero, F. et al. (2014): *PLoS ONE* 9: e85105.
Casteleyn, G. et al. (2010): *PNAS* 107: 12952-57.
D'Alelio, D. et al. (2009): *Protist* 160: 9-20.
Darwin, C. (1859): *On the Origin of Species.*
Davey, J.W. et al. (2011): *Nature Rev. Genet.* 12: 499-510.
De Gelas, K. & De Meester, L. (2005): *Mol. Ecol.* 14: 753-64.
De Meester, L. et al. (2002): *Acta Oecol.* 23: 121-35.
Degerund, M. et al. (2012): *J. Plankton Res.* 34: 416-31.
Drummond, A.J. et al. (2015): *Mol. Biol. Evol.* 22: 1185-92.
Drummond, A.J. et al. (2012): *Mol. Biol. Evol.* 29: 1969-73.
Dvořák, P. et al. (2014): Mol. Ecol., 23: 5538-51.
Dvořák, P. et al. (2015): PNAS, 112: E2559.
Evans, K. M. et al. (2009): *Protist* 160: 386-96.
Fenchel, T. (2005): *Aquat. Microb. Ecol.* 41: 49-54.
Fenchel, T. & Finlay, B.J. (2004): *Bioscience* 54: 777-85.
Fenchel, T. & Finlay, B.J. (2006): *Phil. Trans. R. Soc. B* 361: 1965-73.
Foissner, W. (1999): *Protist* 150: 363-8.
Gächter, E. & Weisse, T. (2006): *Aquat. Microb. Ecol.* 45: 291-300.
Gelman, A. et al. (2003): *Bayesian data analysis*. Boca Raton: Chapman & Hall/CRC.

Gerstein, A.C. & Moore, J.-S. (2011): *Mol. Ecol.* 20: 4385-7.
Hennig, W. (1966): *Phylogenetic Systematics.* University of Illinois Press
Hudson, R.R. (1990): *Gene genealogies and the coalescent process.* Oxford surveys in evolutionary biology, Vol. 7. Oxford Univ. Press.
Hudson, R. R. (2000): *Genetics* 155: 2011-4.
Iglesias-Rodríguez, M. D. et al. (2006): *J. Phycol.* 42: 526-36.
Jensen, J.L. et al. (2005): *BMC Genetics* 6: 13.
Jo, B.Y. et al. (2013): *Phycologia* 52: 266-78.
Keller, E.F. & Lloyd, E.A. (1994) *Keywords in Evolutionary Biology.* Harvard Univ. Press.
Kling, H.J. & Kristiansen, J. (1983): *Nord. J. Bot.* 3: 269-90.
Kooistra, W.H.C.F. et al. (2001): *J. Sea Res.* 46: 213-22.
Kristiansen, J. & Preisig, H.R. (2007): *Süsswasserflora von Mitteleuropa, 1, 2, Synurophyceae.* Spektrum, Heidelberg.
Kynčlová, A. et al. (2010): *Nova Hedwigia Beih.* 136: 283-98.
Lazarus, D. (1983): *Paleobiology* 9: 327-40.
Leigh, J.W. et al. (2011): *Genome Biol. Evol.* 3: 571-87.
Lemmon, A.R. & Lemmon, E.M. (2012): *Syst. Biol.* 61: 745-61.
Librado, P. & Rozas, J. (2009): *Bioinformatics* 25: 1451-2.
Logares, R. et al. (2007): *Microbial Ecol.* 53: 549-61.
Logares, R. et al. (2008): *Environ. Microbiol.* 10:1231-43.
Louis, E.J. (2011): *Fung. Biol. Rev.* 25: 136-42.
Lowe, C.D., Kemp, S.J. & Montagnes, D.J.S. (2005): *Aquat. Microb. Ecol.* 41: 67-77.
Mamanova, L. et al. (210): *Nature Methods* 7: 111-8.
Matvienko, O.M. (1965): *Viznacnik prisnovodnych vodorostej ukrainskoj RSR. 3. Chrysophyta.*
Masclaux, H. et al. (2009): *Limnol. Oceanogr.* 54: 1323-32.
Mayr, E. (1963) *Animal Species and Evolution.* Harvard University Press.
Maddison, W.P. & Maddison, D.R. (2015) http://mesquiteproject.org.
Meyer, M. & Kircher, M. (2010): *Cold Spring Harb. Protoc.* doi:10.1101/pdb.prot5448.
Nylander J.A.A. 2004. MrModeltest v2. Program
Myles, S. (2013): *Trends Genet.* 29: 190-6.
Němcová, Y. (2014): *Nord. J. Bot.* 32: 875-80.
Němcová, Y. et al. (2012): *Nova Hedwigia* 95: 1-24.
Nowack, E.C.M. et al. (2005): *Protist* 156: 239-51.
Péterfi, L.S. & Momeu, L. (2009): *Contrib. Bot.* 44: 107-17.
Pichrtová, M. et al. (2013): *Microb. Ecol.* 65: 68-83.
Plummer, M. (2003): *JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling.* Vienna, Austria.
Poulíčková, A. et al. (2014): *Plos ONE* 9: e86247.
Pritchard, J.K. et al. (2000): *Genetics* 155: 945–59.
Rengefors, K. et al. (2012): *Mol. Ecol.* 21: 3200-9.
Riisberg, I. & Edvardsen, B. (2008): *Eur. J. Phycol.* 43: 413-22.
Ronquist, F. & Huelsenbeck, J.P. (2003): *Bioinformatics* 19: 1572-4.
Rota, C.T. et al. (2011): *Ecography* 34: 659-70.
Rynearson et al. (2006): *Limnol. Oceanogr.* 51: 1249-61.
Ryšánek, D. et al. (2016a): *Phycologia* (in press)
Ryšánek, D. et al. (2016b): *FEMS Microbiol. Ecol.* (in press), http://dx.doi.org/10.1093/femsec/fiw039
Sanderson, M.J. (2003): *Bioinformatics* 19: 301-2.
Sandgren, C.D. & Flanagin, J. (1986): *J. Phycol.* 22: 206-16.
Škaloud, P. et al. (2012): *Phycologia* 51: 303-29.
Škaloud, P. et al. (2013a): *Nord. J. Bot.* 31: 385-402.
Škaloud, P. et al. (2013b): *Nova Hedwigia, Beih.* 142: 141-6.
Škaloud, P. et al. (2014): *Eur. J. Phycol.* 49: 213-29.
Škaloudová, M. & Škaloud, P. (2013): *Phytotaxa* 130: 34-42.
Soberón, J. & Peterson, A.T. (2005): *Biodiv. Inform.* 2: 1-10.
Souffreau, C. et al. (2010): *Phycologia* 49: 309-24.
Tamura, K. et al. (2013): *Mol. Biol. Evol.* 30: 2725-9.
Thuiller W. et al. (2009): *Ecography* 32: 369-73.
van der Gast, C.J. (2015): *Env. Microbiol.* 17: 544-6.
Vyverman W. et al. (2007): *Ecology* 88: 1924-31.
Watts, P.C. et al. (2013): *Biol. Lett.* 9: 20130849.
Wawrzyniak, L.A. & Andersen, R.A. (1985): *Nova Hedwigia* 41: 127-45.
Wiens, J.J. (2011) *Phil. Trans. R. Soc. B* 366: 2336-50.
Wujek, D.E. & Igoe, M.J. (1989): *Nova Hedwigia Beih.* 95: 269-80.