Ecological and genomic consequences of a novel hybrid zone forming in chiffchaff (*Phylloscopus collybita sp.*)

Applicant: Dr. Daria Shipilina
Co-applicant: Prof. Nicholas H. Barton

Contents

1   Introduction and study aims

The aim of this study is to quantify both the ecological and genomic effects of novel hybridization between two subspecies of the common chiffchaff (*Phylloscopus collybita*). This system offers a rare opportunity to observe the beginning of genetic introgression between species in nature, as the range of these subspecies will overlap in the near future, as the collybita subspecies rapidly expands north. The proposed research will lead to a better understanding of the evolution of ecologically relevant traits and their genetic underpinnings during hybrid zone formation. In order to characterize the formation of the hybrid zone and observe natural selection in action we propose:

1. To quantify the effects of habitat usage on established and novel hybrid populations between *P. c. abietinus* and *P. c. collybita*, including monitoring the population dynamics and niche (vegetation) modeling.

2. To explore vocalization changes under the stress of hybridization, using a novel machine learning based approach.

3. To identify targets of selection before and after hybridization in chiffchaff, so as to characterize the genomic consequences of rapid distribution changes that lead to species intermixing.

To this end, we will develop new methods for evolutionary studies. Our new song analysis and recognition tools will be applicable to other study systems where acoustic communication is the key to for species divergence, whilst ecological niche modeling could be widely used in conservation biology, where particular ecotypes are crucial.

A broad range of fundamental questions can be asked within each of proposed aims. We outline them and provide a brief scientific background in the next chapter and also emphasize innovative aspects of the research. Then we justify a choice of scientific methods and provide their detailed explanation for each of the project aims. In the concluding part of the proposal we discuss impact of this project for the receiving institute (IST Austria) and provide a plan for the dissemination of results.

2   Fundamental questions and scientific background

2.1   Following evolutionary processes in a newly-forming secondary contact

Hybrid zones, defined as regions where genetic introgression occurs between two genetically distinct taxa (Barton & Hewit, 1985). They are traditionally in focus of evolutionary biology research as they allow for close observation of evolutionary processes which lead to differentiation between populations and therefore, to speciation (Abbot *et al.*, 2016; Barton & Hewitt, 1985; Hewitt, 2001). Hybrid zones vary both in spatial and temporal aspects and the variation can be caused by strength of selection against hybridization and various factors contributing to partial reproductive isolation (Abbot *et al.*, 2013; Harrison, 1993). Typically, hybrid zone studies have started with observations of phenotypic hybrids in nature. However, quantification of hybridization zone dynamics have been performed long after the initial establishment of the zone. Therefore, exact dating of the initial introgression events and evolutionary consequences of intermixing has rarely been possible to describe (Abbot *et al.*, 2013). The proposed study aims at monitoring the development of a novel secondary contact zone. We will characterize the relative importance of different evolutionary processes in determining the rate of hybridization and their evolutionary consequences for the parental taxa.

In Sweden, a subspecies of the common chiffchaff (*Phylloscopus collybita collybita*) is rapidly expanding northwards (Hansson, 2000; Lindström et al., 2007). In the near future, this subspecies will overlap with the northern form *P. c. abietinus*. Recent data suggest that they already occasionally occur in sympatry in some regions (Hansson, 2000). We assume that the rapid expansion will be followed by hybridization and provide following facts confirming our hypothesis:

1. Within the same pair of subspecies, another hybrid zone in central Europe has been observed (Clement & Helbig, 2008; Cramp, 1992; Ticehurst, 1938). However, it is important to note that the time of isolation in Sweden has been significantly longer. Bird distribution records confirm that there were no chiffchaff of any kind in central Sweden 100 years ago (Lindström *et al.*, 2007).

2. Hybridization between various chiffchaff species is common in general (Bensch *et al.*, 2002; Helbig *et al.*, 2001, Shipilina *et al.*, 2017; Talla *et al.*, 2017).

3. Previous mitochondrial DNA analysis suggests that subspecies *collybita* and *abietinus* have approximately 1% genetic divergence, this translates to a divergence time of several hundred thousand years (Helbig *et al.*, 2001). However, this dating needs to be confirmed with genomic data since analysis of a small set of microsatellite loci found no significant genetic difference between these taxa (Hansson, 2000).

4. Amongst various traits that typically contribute to strong reproductive isolation barriers, only differences in habitat choice appear to be well defined (Hansson, 2000; Lindström *et al.*, 2007). Both subspecies are very similar morphologically (Clement & Helbig, 2008; Cramp, 1992) and have virtually similar vocalization (albeit detailed characterization of vocalizations are currently lacking).

Above we outlined a number of arguments that strongly support that hybridization between two chiffchaff subspecies *will occur* when a zone of secondary contact is formed. However, the *outcome* of such process is an open question, as hybridization can follow different patterns. First, hybridization can slow down species diversification if significant portions of the genome can introgress (Barton & Hewitt, 1985; Taylor *et al.*, 2006). Second, hybridization can serve as a source of adaptive introgression and lead to more rapid adaptive responses in parental lineages (Wu, 2001). Third, hybridization may be deleterious and selection against hybridization may drive the final stage of the speciation process (reinforcement)(Abbot *et al.*, 2013; Servedio & Noor, 2003). In the chiffchaff study system, several outcomes of the secondary contact are possible. If a balance between selection and hybridization is established, and involves multiple genetic loci, we expect to see that only limited genomic regions introgress. Analysis of whole genome data within a hybrid zone is thus a promising approach to improve our understanding of adaptive divergence, and to identify the genomic regions that cause reproductive barriers. With the development of whole genome sequencing methods, we can quantify and characterize genetic differences more precisely. In our recent publications (Shipilina *et al.*, 2017, Talla *et al.*, 2017), we estimated levels of genetic divergence and identified candidate genes which may be under selective pressure in a hybrid zone between *abietinus* and the Siberian chiffchaff species, *P. tristis*. In this study we have a possibility to compare the genetic differentiation landscapes of these two hybrid zones and investigate potential parallel patterns, suggesting that particular traits play important roles for mate recognition and maintenance of reproductive integrity. The aim here is to use a population genomic approach to identify candidate genes related to differences in adaptations and genes involved in vocalization differences.

2.2   Ecology as an isolation barrier

In the common chiffchaff (*Phylloscopus collybita*) subspecies pair, niche preference presumably has the strongest effect on hybridization and is likely to provide strong pre-mating pre-zygotic reproductive isolation (Hansson, 2000), at least in Sweden. However, this effect was never evaluated quantitatively, and many questions remain unresolved. One of the central aims of the present study is to formulate and validate a hypothesis for how habitat choice may shape a hybridization zone formation and future dynamics. Below, we formulate this hypothesis and describe the research methods to be used in its validation.

There is extensive variation in habitat requirements in different parts of the breeding range for each of the subspecies (Price, 1991; Ptushenko, 1954; Ticehurst, 1938). We expect that with such a

variety of ecological niches, some of them could potentially become suitable for both taxa in sympatry, promoting swift intergradation. However, multiple observations in the hybrid zones within the chiffchaff complex (Clement & Helbig, 2008; Hansson, 2000; Helbig *et al.*, 2001) show that habitat choice is narrowed down to one type of plant community or elevation range per subspecies under the pressure of hybridization. For instance, in the proximity of the territory of the expected hybridization zone, *abietinus* prefers old spruce forests, while collybita breeds in deciduous forests (Hansson, 2000).

Motivated by the earlier studies, we put forward a hypothesis that chiffchaffs may show *ecological differentiation* in a newly forming hybrid zone. More specifically, we expect different forest types and vegetation structure to strongly influence the configuration of a sympatric zone.

Testing such hypothesis requires a novel set of methods. After several generations of introgression the population will have a fraction of hybrids and backcrosses. These individuals may occupy intermediate habitats, which thus need a very detailed classification. Therefore, in order to be successful, the methodological approach must satisfy the following criteria:

1. Possibility of detailed classification of preferred plant communities and extraction of community subclasses/syntaxa ("associations", "alliances", "orders")

2. Genetic tools to determine the hybrid index of individual birds

3. Quantification of level of habitat type admixture within each territory

4. Ability to include altitudinal data

We propose a comprehensive approach which combines genome sequence data, remote sensing, and both topographical and vegetation data. In the proposed study, we will perform an in-depth analysis of phytosociology in the hybrid zone using real data on plant species diversity and abundance (Braun-Blanquet, 1964) and remote sensing data from Sentinel 2A/2B satellites. Our previous pilot study in a different species of chiffchaff shows the efficicacy of this approach (Komarova & Shipilina, 2010). We will especially focus on red channel satellite images, which are proven to be the most efficient for identification of borders between plant communities (eg. Teillet *et al.*, 1997; Steven *et al.*, 2003). Another advantage is representation of the data in GIS format, which allows for extensive interpolation of the model to perform the analysis on the geographically wide zone of inbreeding (Hunsaker *et al.*, 2013; Wadsworth & Treweek, 1999).

Population size changes and range expansions might affect competition between diverging lineages on different levels, leading to changes in habitat utilization, migration patterns, and feeding strategies (Burton *et al.*, 2010). Thus, it is important to quantify to understand how diversity can be maintained in a rapidly changing environment.

2.3   Strength of acoustic isolation

Divergent sexual signals, such as bird song, may act as a behavior barriers in sympatry, preventing hybridization (Price, 2008; Hoskin & Higgie, 2010; Tietze *et al.*, 2015). Alternatively, copying of the heterospecific song may result in new acoustic phenotype - "mixed singing" (Catchpole & Slater, 2008; Helb *et al.*, 1985; Jarvis, 2004). In the subspecies pair *collybita-abietinus*, vocalizations are similar (Hansson, 2000; Helbig *et al.*, 2001) but a detailed characterization of acoustic data was not performed. Hence, one of the aims of the present study is to *quantify song variation* within and across subspecies and quantify the potential role of vocalization differences in reproductive isolation.

Our working hypothesis is that acoustic differences between *collybita* and *abietinus* chiffchaffs were previously *underestimated*; the importance of acoustic signals in reproductive isolation requires a careful detailed study. Our assumption is based on the fact that only a few frequency and velocity variables have been measured in subspecies comparisons (Hansson, 2000; Helbig *et al.*, 2001). However, chiffchaff song consists of discrete elements that combine to build phrases, which can be highly diverse between subspecies (Marova *et al.*, 2009; Marova *et al.*, 2013; Salomon & Hemim, 1992; Shipilina *et al.*, 2018). Therefore, within a constant frequency range, song may have pronounced differences in syntax.

We believe that vocal element diversity and usage, as well as phrase composition, are key features in subspecies differentiation (Price, 2008; Hoskin & Higgie, 2010). We will use neural networks to quantify song syntax information and perform time-efficient analysis of large data sets. More specifically, we will use a deep convolutional network for recognition of vocal elements in sonograms (eg. Lasseck, 2013), and devise an algorithm for subspecies identification based on diagnostic/characteristic vocalization patterns.

Another open question is whether acoustic traits contribute to reproductive isolation directly, or differences in vocal behavior are consequences of ecological adaptation. Diversity in bird vocalizations may depend on habitat structure and atmospheric conditions. For example, lower frequency ranges are associated with denser and closer environments (Pearse *et al.*, 2018). In this study, we will determine correlations between habitat choice and acoustic characteristics.

A variety of other questions can be addressed in the future. If song serves as an isolation barrier, what changes in vocalization do we expect to see under the pressure of intergradation? Will we observe formation of a subset of individuals with a "mixed" singing type and how frequent will this song type be? Can we demonstrate reinforcement of acoustical traits (for example, further differentiation in frequency range, specialization in vocal element usage and phase formation in sympatry as compared to allopatry)?

3   Innovative aspects, novelty and impact

Proposed project allows to investigate two fundamental evolutionary processes: process of hybridization and, as a consequence, speciation and formation of reproductive isolation. The present study has two main innovative components:

1. *Study system.* Chiffchaff model provides a unique opportunity to study a formation process of a new hybrid zone in the real time.

2. *Interdisciplinary approach* enabled by synergy between methods of population genetics, machine learning, geobotany, statistical data analysis, and conventional tools used in studies of the ecological problems.

Each of these aspects is described in more details below. Moreover, we expect that our study will also impact environmental and conservation studies.

Study system. Chiffchaff is a representative of the unique group of cryptic species. Species within chiffchaff complex are one of the most difficult groups for field identification (Clement & Helbig, 1998; Svensson & Baker, 1992). In *collybita* and *abietinus* phenotypical identification criteria are even more blurred due to existence of ongoing hybridization in Central Europe. By obtaining large data set, we will clarify true identification criteria for each of the subspecies. We will develop an algorithm allowing subspecies distinction in the field, based on their vocalizations. Further outcome of this study can be used for monitoring and conservation purposes of chiffchaff and other species from a warbler family.

Interdisciplinary approach. We bring together methods which were developed in isolation and, as far as we are aware, were never used before within the same study. For habitat modeling, we will combine classic plant sociology and remote sensing data analysis, which in turn will require developing of simple machine learning algorithm. Even more advanced approaches of machine learning and sound recognition will be applied to bioacoustics data analysis, while preparation of the data sets will be done with a traditional bioacoustics toolbox. Finally, comprehensive population genetic analysis will clarify possible targets of selection and will help to evaluate genetic landscape of forming hybridization zone. Here bioinformatics and mathematical modeling will be combined.

4   Methods and preliminary data

4.1   Work plan and experimental timetable

Applying an integrative approach is essential for the proposed project to be successful. We will combine classical morphological and ecological biometrics with modern bioacoustics analysis, ecological niche-modeling, and large-scale genetic analysis. Work on this project will follow a two-year plan graphically represented in Figure 1. In this section we first describe steps to be taken in preparation for field data collection, and outline our sampling protocol. Further methods are presented in three sections in correspondience to our aims: 1) *evaluation of habitat preferences*, 2) *bioacoustics analysis*, and 3) *genomics*. Throughout, we also demonstrate how our preliminary data will enrich the proposed forthcoming analyses. We conclude by evaluating of ethical aspects of the proposed research.

figpage2.pdf

4.2 Field work: transects and sampling locations

The main focus of our sampling efforts is the currently developing hybrid zone in central Sweden. Additionally, data from the 'old' Central European hybrid zone will be obtained for comparative purposes. To allow for further clinal analysis and to ensure that we capture entire gradients of morphological, vocal, and genetic features, data collection will be performed along extended transects (for preliminary locations, see Figs. 2, 3). Two transects will be sampled during field seasons of 2019 and 2020 (Fig. 1: steps 3, 10). One will span the currently forming hybrid zone in Sweden, starting in the center of the distribution of allopatric *P. c. collybita* in central Europe (Austria) and ending in the center of the distribution of allopatric *P. c. abietinus* in NE Sweden. Two parts of this transect are shown on Fig. 2 and 3. The second transect will cover the established hybrid zone in Eastern Europe, again starting from *P. c. collybita* allopatric regions in Austria and finishing in eastern Poland (allopatric *P. c. abietinus*).

To access fine structure of allopatric populations we will sample individuals from different parts of the subspecies range (Austria, Spain - *collybita*, Sweden, Finland – *abietinus*). The aim is to have a total dataset of 30 individuals from each of the allopatric zones and 120 individuals from within the hybrid zones along each of the transects (300 samples in total). Sampling permits will be obtained in the countries of proposed sampling after transect refinement.

Identification of chiffchaff subspecies in the field requires experience. As mentioned above, songs of *collybita* and *abietinus* are not easily distinguishable by ear (Svensson & Baker, 1992). Therefore, we will develop a machine learning algorithm for subspecies recognition in preparation for the first field season. This will allow us to distinguish subspecies during sampling (see Aim 2, acoustics section for a description of the proposed recognition algorithm).

4.3 Data collection

All field work will be done in early summer when territories have been firmly established and seasonal migration has come to an end. We will follow the previously described transects (see previous section). We will make stops approximately every 50-100 km for sampling of up to 10 male individuals. Previously developed sampling protocols, which include song recording, capturing, measurements and DNA sampling will be used (Shipilina *et al.*, 2017). Chiffchaffs are released right after sampling. The main equipment is already available in the host laboratory. Sampling will be done mainly by Daria Shipilina with assistance of collaborators and volunteers.

We will use the following sampling protocol (please see ethical note on reducing possible impact on animals):

1. Song will be recorded for ten minutes (approximately 50 individual songs (analysis units)will be present). Necessary equipment (recorder, directional microphone) is available through collaboration with Prof. Niclas Backström (Uppsala University).

2. After song recording, male birds will be captured using a standard sound trap (playback and mist netting) approach. Diverse biometric data will be collected to compare morphology between subspecies with special focus on wing length, tail length, body mass and pointedness of the wing index (Salomon & Hemim, 1992). Quantification of plumage coloration will be done by estimating of intensity of yellow (lipochrome) coloration on throat, breast, belly, and upperparts of the body (Svensson, 1992; Ticehurst, 1938).

3. Finally, a blood sample will be collected using standard procedures (puncture of the wing vein and collection of a small amount of blood using glass capillaries). Blood will be stored in Queens lysis buffer to facilitate extraction of high quality DNA for analysis at a later stage.

4. After sampling, chiffchaffs will be immediately released. In our experience (Shipilina *et al.*, 2017) individuals returned to their regular activities (feeding, advertising a territory) right after release.

5. In the core of the hybrid zone we will perform an in-depth analysis of phytosociology following classic Braun-Blanquet protocol (Braun-Blanquet, 1964). Data on plant species diversity and abundance will be collected on a 10x10 m^2 square plots (or relevés) in the core of individual territory (one plot per individual for at least 20 individuals from each of the contact zones).

4.4   Methods for Aim 1: Plant ecology

Quantification of differences in habitat choice in chiffchaff will be performed in two steps: collection of the field data (Fig.1: steps 3,10) and analysis of remote sensing data in the lab (Fig.1: steps 4-5, 11).

During the field season we will collect data on plant species in the overlap zone following standard protocols (n=20)(Braun-Blanquet, 1964). On the individual chiffchaff territory (i.e., where a male is actively and continuously singing) data on present plant species, their abundance and developmental stage will be collected. Further analysis will be implemented in collaboration with Dr. Anna Komarova. At the next step, communities will be classified (software package TWINSPAN; Roleček *et al.*, 2009) and divided into separate associations (the smallest plant community unit). Associations will be grouped into a larger ecological conceptual units ("alliances", and further to "orders" and "classes") to build an hierarchal system of described communities. We expect that habitat preferences between subspecies will be visible on the level of bigger classification units - alliances (eg. deciduous vs. coniferous forests). Distinction between smaller units - associations - will help to clarify niche preferences between hybrids and parental forms.

Next we will use remote sensing data (images from multiple channels) from the Sentinel satellite, which has an exceptional resolution of 20 m. Sentinel images are open for public use and can be freely used for research purposes. A supervised learning algorithm (Random Forest) will be used to classify associations on the images. GPS coordinates of individual chiffchaff territories with classified plant associations will be used as a training set. As a result of this step, we will have each of the individuals, inhabiting both the allopatric regions and the hybrid zones, assigned to plant community units (associations, alliances) and can develop a model for further prediction of chiffchaff preferences in other breeding areas.

At the final stage we will apply multiple statistical tests to check for correlation between subspecies identification and habitat choice. Additionally, we expect several individuals falling into an intermediate classes due to their hybrid origin, especially in the old hybrid zone.

4.5   Methods for Aim 2: Bioacoustics research methodology

Evaluation of acoustic differences of chiffchaff songs will be performed in two steps. First, at the preparation stage of the project, we will develop a preliminary algorithm allowing us to distinguish between *collybita* and *abietinus* subspecies in the field (Fig. 1: step 2). This will be done using already available song data recorded in the allopatric regions in 2018. Next, we will refine the algorithm using data collected in the field and estimate levels of song admixture in the hybrid zone (Fig. 1: step 6).

Preliminary algorithm. Subspecies song recognition algorithm will be developed in close collaboration with Dr. Ekaterina Putintseva (IST Austria). We will apply methods of image recognition to spectrograms (graphical representation of frequency from time) from both subspecies. Data for our training sets will be obtained from open source (60 *collybita*, 41 *abietinus* individuals from xeno-canto.org, 57 individuals from macaulaylibrary.org) and our own collections (Eastern Pyrenees, 24 individuals recorded in 2018). We will filter the data for noise and quality, then test various software packages (RavenPro: Charif *et al.*, 2009; warbleR: Araya-Salas & Smith-Vidaurre, 2017; PRAAT: Boersma & Weenink, 2010) to develop the most efficient pipeline for automated extraction of spectrograms/images of individual songs. We will make snapshots of the spectrogram of the standard length of 7 seconds (examples on the Fig.4). Daria Shipilina will be responsible for data preparation.

In the next step, we will apply deep convolutional neural networks (Waseem & Zenghui, 2017) and recurrent neural networks for subspecies identification based on vocalizations. The application of
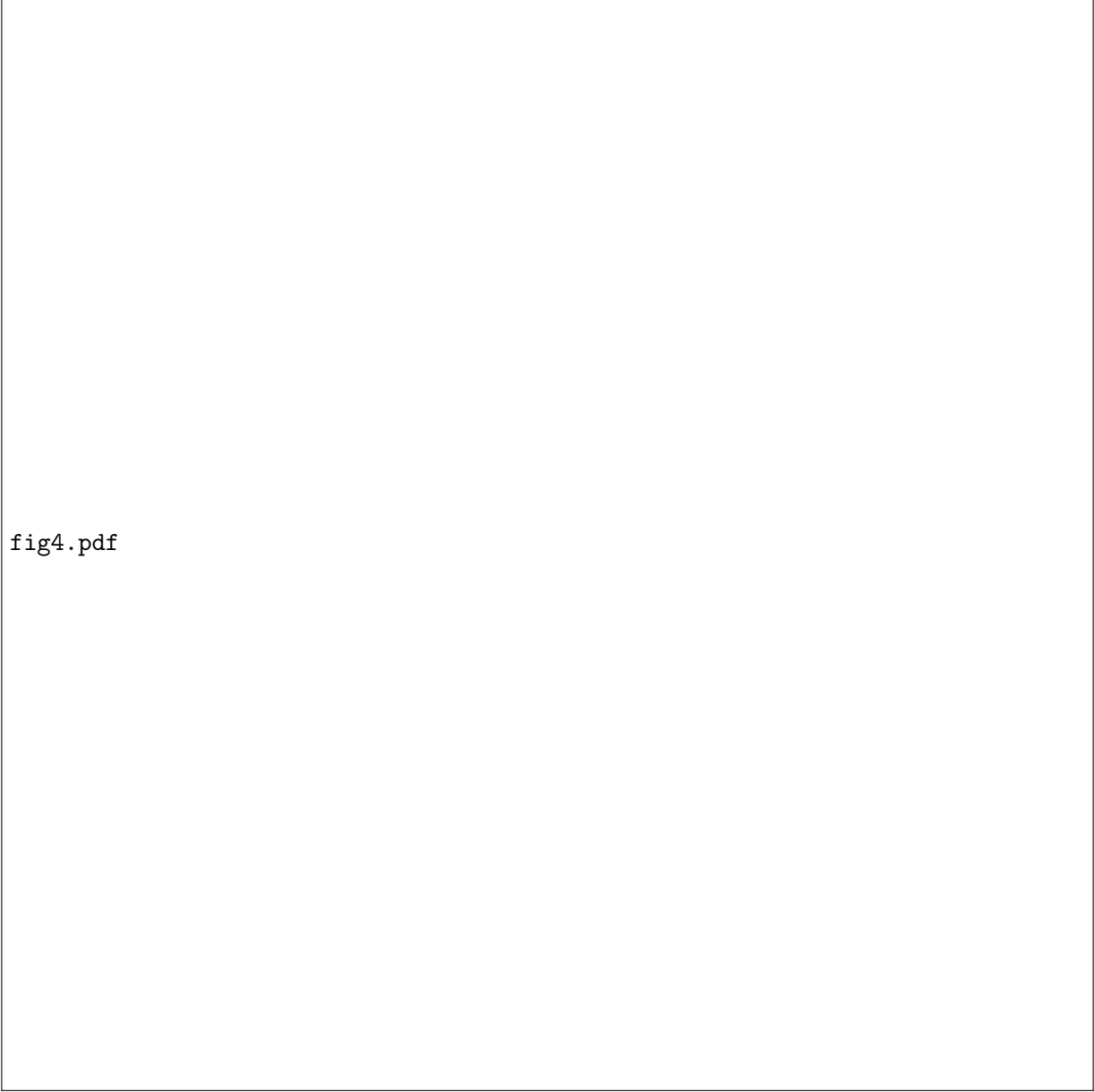
fig4.pdf

Figure 1: Songs of allopatric collybita and abietinus chiffchaffs. Spectrogram snapshot, which will be used as training set: a) France, recording by F.Deroussen (xeno-canto.org), b) Eastern Pyrenees, D.Shipilina, c) Russia, D.Shipilina, d) Sweden, C.Brinkman (xeno-canto.org).

recurrent neural networks is a classical machine learning approach for speech recognition (Graves *et al.*, 2013; Mandic & Chambers 2001), while convolutional neural networks are increasingly used for sound recognition discrimination (for example, Hershey *et al.*, 2017, Kumar, 2017). We will optimize a number of different ensemble architectures, either based on initial recognition of vocal elements with a subsequent classification, or a one-step analysis of a whole spectrogram of individual songs. Despite the wide application of machine learning algorithms in bird recognition, identification at the subspecies/individual level is still rare. We foresee a wider range of applications for our algorithm in the scientific community and also for bird enthusiasts in general.

Population analysis. Song element diversity and usage, as well as phrase composition, will be key features in subspecies differentiation. At the next step, we will use attention-based deep neural networks (Xu *et al.*, 2015) in order to to determine the most distinct acoustic features of the two subspecies and quantify vocal admixture levels in the hybrid zone. Measurements of song admixture level along the gradient of hybridization will be applied to acoustical cline analysis and later compared to genetic clines.

4.6   Methods for Aim 3: Population genetics

On the genomic level, the most straightforward task with immediate effects in the interpretation of how hybridization will affect the evolutionary fate of the two subspecies, is to find barriers to introgression in the newly forming secondary contact zone. By adding genomic data from an established hybrid zone, I will be able to determine introgression rates across chiffchaff genomes, identify potential barriers to gene-flow and characterize the specific characteristics (genes, functional non-coding elements) of such regions (see eg. Feder *et al.*, 2017; Westram & Ravinet, 2017). To our knowledge, such a study in combination with quantification of ecological and behavioral traits has not yet been done.

At the preliminary stage of genomic analysis, I will evaluate the level of genetic differentiation between subspecies by deep individual sequencing of 30 *P. c. collybita* and 30 *P. c. abietinus* (20 new, 10 available from previous studies) individuals from allopatric populations (see Methods for data collection). We expect sufficient level of genetic differentiation within and between subspecies (Helbig *et al.*, 2001). Basic population genetic summary statistics (Korneliussen *et al.*, 2014) will be calculated in non-overlapping windows and anchored on our previously established chiffchaff reference genome (Shipilina *et al.*, 2017; Talla *et al.*, 20017). Based on these data, we will:

1. Demonstrate how rapid introgression shapes genomic composition of the forming hybrid population. In high-coverage sequencing data from this population (at least 30 individuals) I expect to see elevated linkage disequilibrium and will use this to detect recent exchange of large genomic blocks from the parental forms (see e.g. Kong *et al.*, 2008). Additionally, I will evaluate genomic regions of reduced diversity and elevated divergence, characterizing their associations with particular genes.

2. Compare levels of genetic admixture between new and established hybrid zones. I will infer geographic clines using allele frequency shifts (Barton & Hewitt, 1985; Gompert & Buerkle, 2012). As a preliminary step, I will develop a panel of SNPs, which will allow for a cost-efficient sequencing of an extra 250 individuals from both hybridization zones. Further, I will compare allele frequency clines with clines of acoustic and morphological features, synthesizing the main approaches of my proposal. This will allow for discriminating between ecology, behavior, and random effects in driving subspecies divergence.

3. Model the history of the existing hybrid zone and compare it to observation of new hybrid zone formation, using coalescent based methods (Lohse *et al.*, 2016).

4.7   Ethical note

Chiffchaff are abundant in the area of research and have a "least concern" status (LC, The IUCN Red List of Threatened Species). The field work will follow sampling protocols, which we specifically

developed for *Phylloscopus* species and optimized to reduce impact on the animals. Capturing of the chiffchaff will be implemented with a "song trap". First, conspecific song will be played to attract an animal (around 10 minutes of exposure). The attracted bird will be trapped in a mist net, but immediately released, measured (wing and tail length) and photographed. A blood sample for DNA analysis will be taken using a standard technique for the wing vein (Arctander, 1988). Total handling time will not exceed 30 minutes and after that the bird will be released at the same spot as it was captured.

Sampling in nature requires permits (in concordance with the Nagoya protocol) from each country where field work is planned. In the case of success of this proposal, the corresponding sampling permits will be obtained.

5   Research facility, scientific collaborations

IST Austria will be an excellent platform to conduct this research project. Excellence is promoted by working with Prof. Nicholas Barton and supported by state of the art research facilities of the institute and wide interdisciplinary collaborations. I believe, that I can contribute to the excellence on the IST by bringing a new study object, together with expertise in ecology, bioacoustics, and bioinformatics.

Professor Nicholas Barton is a leading expert in population genetics and a well-renowned author of both theoretical and empirical studies on hybrid zone dynamics. Prof. Barton studies theoretical aspects of various evolutionary concepts and continuously develops new quantitative methods allowing for testing the newest theories against empirical data. Working in the Barton lab at IST in Austria guarantees access to the best possible theoretical foundation for understanding hybrid zone dynamics and gives me an opportunity to contribute to development of those approaches by testing them on a new study object and providing bioinformatics support with preparation of genomic data for modeling. In addition to abstract computation, Barton group has broad experience in field studies of the hybrid zone (*Antirrhinum* project) and expertise in plant community and conservation ecology. The wide research interests of the Barton lab will promote my professional development, deepen my understanding of evolutionary processes and will provide a creative atmosphere for synthesis of major unresolved questions in evolutionary biology.

IST Austria is a vibrant research venue with many national and international collaborators at the very forefront of evolutionary biology research (ex. Prof. Fyodor Kondrashov, Prof. Beatriz Vicoso groups). Additionally, IST hosts multiple internationally renowned researchers, both in the fields of ecology and genomics, and provides an immense potential for interaction across departments and research groups and collaboration on many aspects of the project, not the least computer science. In particular, an interdisciplinary collaboration with **Dr. Ekaterina Putintseva** within IST will allow me to apply various machine learning algorithms for the advanced bioacoustics analyses. Her research at IST is focused on applying machine learning to multiple evolutionary questions. Additionally, we will have an opportunity to benefit from input from other computer science groups at IST: groups of Prof. Christoph Lampert and Prof. Dan-Adrian Alistarh.

IST Austria has shared access to multiple facilities that are essential for my project. IST Life Science facilities provide a platform for swift analysis of high quality next-generation sequencing data and access to a large and well-maintained computer cluster necessary for such analysis and efficient neural network algorithm application.

This project methodology and applications will be enriched by continuation of two long-term international collaborations. A continuous collaboration with **Prof. Niclas Backstöm** will lead to novel data for the understanding of the evolutionary past of the entire chiffchaff species complex. Prof. Backstrom is a leading expert in speciation and population genomics, and he works extensively with empirical data on various bird and butterfly species, including chiffchaff. Dr. Backström's input on this project has a significant value for both efficient data analysis and informed interpretation of the results. Moreover, his solid knowledge about Swedish bird population dynamics will be applied both at the planning stage of the project and during field work in the novel hybrid zone.

Dr. Anna Komarova will participate in analysis of subspecies habitat preferences. Her expertise

is plant sociology, habitat modeling and extensive experience with remote sensing analysis will enrich ecological part of the project. Our previous collaborative project (Komarova & Shipilina, 2010) in Russian North (Arkhangelsk region) revealed intriguing patterns in chiffchaff habitat choice and distribution. We will now extend it to larger data set and different subspecies. As researcher at Greenpeace Russia, Anna will consult me on such environmental questions as a forest cover loss in the area of study, therefore, deepening an environmental impact of this study.

6 Strategies for dissemination of results and outreach

Dissemination of the results will be preformed through publication of scientific papers and presentation of our results to scientific community in Austria and internationally. Additionally, results of our studies will be communicated to broader audience through outreach events, press releases, and publications in general-reader scientific journals.

We plan to present our results both within Austrian (for example EvoVienna Meetings) and international scientific communities both in fields of evolutionary biology (conferences of European Society of Evolutionary Biology, the Society for the Study of Evolution) and ornithology (European Ornithological Union Conference, International Ornithological Congress).

In addition to scientific community, our findings will be of an interest of a wide audience of bird enthusiasts and bird conservation organizations. Currently, there is a lot of interest in chiffchaff subspecies identification and distribution (Collinson et al., 2018, Lewis et al., 2018). To make our findings more accessible to a wide audience we plan publication in bird watcher's journals, such as British Birds.

Results of our findings will be presented to a very diverse group of all ages through IST Austria's annual outreach event - Open Campus. This event attracts nearly 2000 visitors each year, we hope that our study will inspire future scientists.

6.1 List of abbreviations

GIS – Geographic Information System
GPS – Global Positioning System
SNP – Single Nucleotide Polymorphism

7 References

1. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, et al. (2013) Hybridization and speciation. J. Evol. Biol. 26: 229-246

2. Abbott RJ, Barton NH, Good JM. (2016) Genomics of hybridization and its evolutionary consequences. Molecular Ecology 25: 2325-2332

3. Araya-Salas M, Smith-Vidaurre G. (2017) warbleR: an r package to streamline analysis of animal acoustic signals. Methods in Ecology and Evolution 8(2): 184–191

4. Barton NH, Hewitt GM. (1985) Analysis of Hybrid Zones. Annual Review of Ecology and Systematics 16: 113-148

5. Bensch S, Helbig AJ, Salomon M, Seibold I. (2002) Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. Molecular Ecology 11: 473–481

6. Boersma P, Weenink D. (2010) Praat: doing phonetics by computer. Retrieved from http://www.praat.org/

7. Braun-Blanquet J. (1964) Pflanzensoziologie, Grundzüge der Vegetationskunde (3rd Edition). Springer-Verlag, Berlin, 631

8. Burton OJ, Phillips BL, Travis JM. (2010) Trade-offs and the evolution of life-histories during range expansion. Ecology letters13(10): 1210-1220

9. Catchpole CK, Slater PJB. (2008) Bird song: biological themes and variations (2nd edition). Cambridge University Press, Cambridge

10. Charif RA, Strickman LM, Waack AM. (2010) Raven Pro 1.4 User's Manual. The Cornell Lab of Ornithology, Ithaca, NY

11. Clement P, Helbig AT. (1998) Taxonomy and identification of chiffchaffs in the western Palearctic. British Birds 91(9): 361–376

12. Cramp S. (1992) Handbook of the birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic, v. 6. Warblers. (S. Cramp, Ed.). Oxford University Press.

13. Collinson JM, Murcia A, Ladeira G, Dewars K, Roberts F, Shannon T. (2018) Siberian and scandinavian chiffchaffs in Britain and Irelang - a genetic study. British Birds 111: 384-394

14. Feder JL, Nosil P, Gompert Z, Flaxman SM, Schilling MP. (2017) Barnacles, barrier loci and the systematic building of species. J. Evol. Biol.30:1494-1497

15. Gompert Z, Buerkle CA. (2012) bgc: Software for Bayesian estimation of genomic clines. Molecular Ecology Resources 12(6): 1168–1176

16. Graves, A., Mohamed, A. R., Hinton, G. (2013) Speech recognition with deep recurrent neural networks. In Acoustics, speech and signal processing (icassp): 6645-6649

17. Harrison RG (Ed.). (1993). Hybrid zones and the evolutionary process. Oxford University Press on Demand.

18. Hansson MC, Bensch S, Brännström O. (2000). Range expansion and the possibility of an emerging contact zone between two subspecies of Chiffchaff Phylloscopus collybita ssp. Journal of Avian Biology 31(4): 548–558

19. Helb HW, Dowsett-Lemaire F, Bergmann HH, Conrads K. (1985) Mixed Singing in European Songbirds — a Review. Zeitschrift Für Tierpsychologie 69(1): 27-41

20. Helbig AJ, Salomon M, Bensch S, Seibold I. (2001). Male-biased gene flow across an avian hybrid zone: Evidence from mitochondrial and microsatellite DNA. Journal of Evolutionary Biology 14: 277–287

21. Hershey S, Chaudhuri S, Ellis DP, Gemmeke JF, Jansen A, Moore RC, Slaney M. (2017) CNN architectures for large-scale audio classification. In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on (pp. 131-135). IEEE.

22. Hewitt GM. (2001) Speciation, hybrid zones and phylogeography - Or seeing genes in space and time. Molecular Ecology 10(3): 537–549.

23. Hunsaker CT, Goodchild MF, Friedl MA, Case TJ (Eds.). (2013) Spatial uncertainty in ecology: implications for remote sensing and GIS applications. Springer Science & Business Media.

24. Hoskin CJ, Higgie M. (2010) Speciation via species interactions: the divergence of mating traits within species. Ecology Letters 13: 409-420

25. Jarvis ED (2004). Brains and birdsong. Roberts & Company Publishers

26. Komarova AF, Shipilina DA. (2010) Habitat choice in the hybrid population of Siberian and Easter European chiffchaffs in Pinega State Reserve. Current Problems of Ecology and Natural Resourses (RUDN) 12: 124-127 (In Russian)

27. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI, Ingason A, Steinberg S, Rafnar T. (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. Nature Genetics 40: 1068–1075

28. Korneliussen TS, Albrechtsen A, Nielsen R. (2014) ANGSD: Analysis of next generation sequencing data. Bioinfo. 15: 356

29. Kumar A, Raj B. (2017) Deep cnn framework for audio event recognition using weakly labeled web data. arXiv preprint arXiv:1707.02530.

30. Lasseck M. (2013) Bird song classification in field recordings: winning solution for NIPS4B 2013 competition. In: Glotin H, LeCun Y, Artieres T, Mallat S, Tchernichovski O, Halkias X, editors. Proceedings of Neural Information Processing Scaled for Bioacoustics:176–181

31. Lewis M, Penn A, Collinson JM. (2018) Subspecies identification in common chiffchaffs wintering at Nigg Bay, North-east Scotland, in 2016/17. British Birds 111: 395-401

32. Lindström Å, Svensson S, Green M, Ottvall R. (2007) Distribution and population changes of two subspecies of Chiffchaff Phylloscopus collybita in Sweden. Ornis Svecica 1984: 137–147

33. Lohse K, Chmelik M, Martin SH, Barton NH. (2016) Efficient strategies for calculating blockwise likelihoods under the coalescent. Genetics 202: 775-786.

34. Nosil, P., Harmon, L. J., Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. Trends in Ecology & Evolution, 24, 145–156.

35. Mandic, D. P., Chambers, J. (2001). Recurrent neural networks for prediction: learning algorithms, architectures and stability. John Wiley & Sons, Inc.

36. Marova I, Fedorov V, Shipilina D, Alekseev V. (2009) Genetic and Vocal Differentiation in Hybrid Zones of Passerine Birds: Siberian and European Chiffchaffs (Phylloscopus [collybita] tristis and Ph. [c.] abietinus) in the Southern Urals. Doklady Biologich. Nauk 427: 384–386

37. Marova I, Shipilina D*, Fedorov V, Ivanitskii V. (2013) Siberian and East European chiffchaffs: geographical distribution, morphological features, vocalization, phenomenon of mixed singing, and evidences of hybridization in sympatry zone. In El. Mosquitero ibérico. Grupo Ibérico de Anillamiento: 119-139

38. PriceT. (1991) Morphology and Ecology of Breeding Warblers Along an Altitudinal Gradient in Kashmir, India. The Journal of Animal Ecology 60(2): 643

39. Price TD. (2008) Speciation in birds. Greenwood Village, CO: Roberts & Company Publishers.

40. Ptuschenko, E. S. (1954). Familia Sylviidae. In G. P. Dementjev & N. A. Gladkov (Eds.), Birds of the Soviet Union, v.6 (pp. 142–398). Moscow: Sovetskaja Nauka.

41. Salomon M, Hemim Y. (1992) Song Variation in the Chiffchaffs (Phylloscopus collybita) of the Western Pyrenees — the Contact Zone between the collybita and brehmii Forms. Ethology, 92: 265-282

42. Roleček J, Tichý L, Zelený D, Chytrý M. (2009) Modified TWINSPAN classification in which the hierarchy respects cluster heterogeneity. Journal of Vegetation Science 20: 596-602

43. Servedio MR, Noor MA (2003) The role of reinforcement in speciation: theory and data. Annual Review of Ecology, Evolution, and Systematics 34(1): 339-364

44. Shipilina D, Serbyn M, Ivanitskii V, Marova I, Backström N. (2017) Patterns of genetic, pheno-typic and acoustic variation across a chiffchaff (Phylloscopus collybita abieti- nus/tristis) hybrid zone. Ecology and Evolution 7: 2169–2180

45. Steven MD, Malthus TJ, Baret F, Xu H, Chopping MJ. (2003) Intercalibration of vegetation indices from different sensor systems. Remote Sensing of Environment 88 (4): 412-422

46. Svensson L, Baker K. (1992). Identification guide to European passerines. BTO 1992.

47. Talla V, Kalsoom F, Shipilina D, Marova I, and Backström N. (2017) Heterogeneous patterns of genetic diversity and differentiation in European and Siberian chiffchaff (Phylloscopus collybita abietinus / P. tristis). G3 (7): 3983-3998

48. Taylor EB, Boughman JW, Groenenboom M, Sniatynski M, Schluter D, Gow JL. (2006) Speci-ation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. Molecular Ecology 15(2): 343-355

49. Teillet PM, Staenz K, Williams DJ. (1997) Effects of spectral, spatial, and radiometric character-istics on remote sensing vegetation indices of forested regions. Remote Sensing of Environment 61(1): 139–149

50. Ticehurst CB. (1938) A systematic review of the genus Phylloscopus. London: Trustees of the British Museum.

51. Tietze DT, Martens J, Fischer BS, Sun YH, Klussmann-Kolb A, Päckert M. (2015) Evolution of leaf warbler songs (Aves: Phylloscopidae). Ecology and Evolution 5(3): 781–798

52. Wadsworth, R., Treweek, J. (1999). GIS for Ecology. Addison Wesley Longman.

53. Waseem R, Zenghui W. (2017) Deep convolutional neural networks for image classification: a comprehensive review. Neural Computation 29(9): 2352-2449

54. Westram AM, Ravinet M. (2017) Land ahoy? Navigating the genomic landscape of speciation while avoiding shipwreck. J. Evol. Biol. 30:1522-1525

55. Feder JL, Nosil P, Gompert Z, Flaxman SM, Schilling MP. (2017) Barnacles, barrier loci and the systematic building of species. J. Evol. Biol.30:1494-1497

56. Wu CI. (2001). The genic view of the process of speciation. Journal of Evolutionary Biology 14(6): 851-865

57. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Bengio Y. (2015) Show, attend and tell: Neural image caption generation with visual attention. In International conference on machine learning (pp. 2048-2057).

8   Applicants CV and list of publications

a

i

9 Co-applicants CV and list of publications

u

f

a