1. Purpose of the visit.

My PhD. project tries to evaluate the effects of the loss of genetic diversity and accumulation of inbreeding on the survival and reproduction of the Iberian lynx, the most endangered felid in the world. The loss of diversity has been previously documented in the species through the use of neutral molecular markers, leaving open the question to what extent patterns for neutral markers represent adaptive variation, given the possible effects of selection on the latter. The project's goal is to address this question using immune-related genes as markers for functional variation.

The aim of my visit to Pr. Sommer's lab was to learn how to make the best use of NGS technologies in my research. Mainly, learning how to properly analyse and interpret 454 data we had already obtained for MHC class I and II genes. As there is not a gold-standard methodology yet to do so I wanted to visit a group that has worked extensively with this type of data and that is designing a new methodology, so I could fully comprehend the whole process. Besides, I wanted to take advantage of their expertise on wildlife immunogenetics to ask for some advice on other candidate genes that that might be most relevant for my study.

Finally, I ment to use the help of their expert bioinformatician Dr. Mazzoni, permanent member of Berlin Center for Genomics in Biodiversity Research, to get some hints about how could I fully exploit the available genomic resources (e.g. an annotated domestic cat genome) and those that are becoming available as the Iberian lynx genome sequencing project progresses (i.e. lynx whole genome and transcriptome assemblies), to select markers and to design assays to assess their variation.

2. Description of the work carried out during the visit.

The work has been very centered on the analysis of the 454 data I already had from Iberian lynx MHC class I and II. As Dr. Mazzoni is designing a new pipeline to analyze this type of data aiming to: i) get more and less biased sequences from the raw 454 sff files, ii) process the reads, iii) call the alleles having less false and missed alleles. During my visit we have been able to fully analyze level my dataset at a populational following Mazzoni's pipeline.

We focused on the analysis of my data to:

a) Ensure that we had high quality data so we could conveniently perform the analysis.
b) Perform the prepossessing analysis for the designed pipeline.
c) Fix the scripts on the pipeline (regarding sequence analysis and database construction) so they fitted my specific requirements.
d) Constructed postgreSQL databases with all the information about the putative alleles to easily interpret the output, filter errors and asses amplification bias.
e) Check each putative allele one by one visually to decide if they were or not an error from a more frequent haplotype and, if it was the case, which kind of error it was.

Regarding outcomes we have:

f) Called the alleles present in the remaining population of iberian lynx.
g) Selected which samples needed a replicate to have a robust allele calling and probably confirm the

existence of some new rare alleles.

h) Selected the appropriate methodology for genotyping a large number of individuals, now that we have an idea about the number of alleles present in the population the number of loci and the amplification-bias of our primers.

Besides, during my meeting with Prof. Sommer she gave me advice on how could we try to relate the measures of genetic diversity in MHC genes with individual fitness.

About the work on different genes, Toll-Like Receptors (TLR) were the next candidate genes to study. During my visit I have learnt which methods were more efficient to find TLR on transcriptomic data available and how to annotate them. TLR transcripts found have been annotated through homology to their best match in SwissProt (using revised aminoacid sequences).

Now we will use this information to carry out a preliminary study on the genetic variation of this gene-family in the 11 individuals sequenced in the genome project to elucidate whether it is an interesting gene-family or not (in terms if genetic variation) to address our main questions.

3. Description of the main results obtained.

At a populational level we have found 13 MHC class I alleles and 5 corresponding to MHC class II. Those figures could increase once the individuals with putative alleles seen just in one PCR are repeated.

Crossing this information with the transcriptomic data we have been able to confirm the expression for 5 and _ alleles in MHC class I and II respectively. Moreover, once we had the real alleles present in the individuals with transcriptome data, we could identifiy errors during the transcript assembly. The rest of the alleles are not assessed yet in terms of expression, as the individuals used to generate the transcriptome do not carry those alleles. Their expression will be latter checked using RT-PCR.

Furthermore, we have been able to identify those alleles that were not efficiently amplified, and some further analysis will help to determine which strategy to use to reduce the bias detected with the amplification methodology currently used. This way we will need less coverage to genotype each individual and thus we will be able to genotype more of them with the same costs.

The work on TLR genes have revealed the TLR transcripts that we had (TLR 1 to 10).

4. Future collaboration with host institution (if applicable)

I will inform Dr. Mazzoni about the final results of my analysis so she can test her pipeline against an input data-set as mine that is very challenging do to the high amplification bias.

We have not formally arranged any future collaboration but we keep the door open to future collaborations.

5. Projected publications / articles resulting or to result from the grant (ESF must be acknowledged in publications resulting from the grantee's work in relation with the grant).

A projected article about MHC class I and II contemporary variation willinclude the outcome of this vist.

6. **Other comments (if any).**