

Grant's gazelle project overview

Contact person: Genís Garcia-Erill genis.erill@bio.ku.dk or Hans R. Siegismund, hsiegismund@bio.ku.dk

Data description

The data set you will be given comes from 91 individuals of Grant's gazelle (*Nanger granti*) and 4 individuals of Thomson's gazelle (*Eudorcas thomsonii*). Grant's gazelle is interesting because it was recently discovered by genetic methods that it probably consists of three species (*N. granti*, *N. notate* and *N. petersii*) rather than one. Some morphological evidence suggests that *N. granti* may be further subdivided into *N. g. granti* and *N. g. robertsii*. The puzzling thing is how strong genetic barriers have evolved without any obvious isolating mechanism, i.e. no landscape barriers to gene flow and no obvious ecological differentiation between the three species.



The data was generated by a technique called RAD sequencing (Restriction site Associated DNA sequencing, or RADseq. RADseq subsamples the genome with a restriction enzyme so that you end up with about 10 megabases of the genome sequenced in each individual. The data set you will be getting consists of PLINK files with SNP genotypes from the variable positions in these 10 megabases. PLINK is a standard SNP format that can be used in many other programs and can be easily converted to other file formats. Remember that the PLINK files represent the full data set, meaning all individuals from both species. For some analyses it will be obvious to subset the PLINK files by species or populations. It could be a good idea to filter the data when you do various analyses, particularly for missing data. The full data set has been pre-filtered removing sites with more than 20% missing data across individuals.

Project data

Your project data (and additional information) is found in the folder

~/groupdirs/SCIENCE-BIO-Popgen_Course/projects/grants/

The file "grants_popInfo.txt" provides a sample information sheet specifying the geographic origin of each sample and its official subspecies designation.

Problems to address

Although you are free to explore any aspect of the data you may like, some of the obvious questions to address are:

- 1) Can you confirm the strong genetic differentiation within the Grant's gazelles, and what are the most prominent groupings (or genetic clusters) within the species complex?
- 2) Is there any indication of admixed populations, i.e. populations that could be hybrids between two of the Grant's species?
- 3) What is the level of genetic diversity in different populations, and does it differ between populations?
- 4) Other possible directions: demographic history and technical/methodological issues.

You are strongly encouraged to familiarize yourself with the literature on these topics. There are at least 3 papers on Grant's gazelle genetics that are obvious to refer to. Keep in mind that the data you will get will be much more comprehensive than what has been published before, so you should be able to get clearer and more detailed results. Remember, there should be a biological interpretation of your results. This could relate to evolutionary processes in general, phylogeography, conservation or something different. Up to you.

Suggested software

This is a non-exhaustive list of computer programs that could be relevant. Use this as a guideline; if you want to do something else or use other programs to do similar analyses, feel free to do so. You are also encouraged to use R to manually (or by using other popgen packages) to explore the data as you see fit. Please also refer to the exercises you have been doing during the course, as there will be methods, R code etc. that you can use in the analysis. Software manuals, examples etc. should be readily available by googling the software names.

- adegenet: an R package that can do basic calculations like heterozygosity, HWE, F_{st} etc. It can also do more fancy "spatial genetics".
- snpmatrix: an R package that can perform PCA. It is pre-installed in the popgen server.
- admixture: a program to infer the admixture proportions under different values of k , the number of assumed clusters. Can be used to find genetic structure. It is pre-installed in the popgen server.
- treemix: a program to infer the drift tree between different pre-determined genetic groups (or populations). It is pre-installed in the popgen server.

Suggested literature

- Siegismund, H.R., E.D. Lorenzen & P. Arctander 2013. *Nanger (granti)* Grant's gazelle species group; pp. 373–379 in Kingdon, J. & Hoffmann, M. (eds). 2013 Mammals of Africa: Volume VI. Pigs, Hippopotamuses, Chevrotain, Giraffes, Deer and Bovids. Bloomsbury Publishing, London.
- Lorenzen, E.D., P. Arctander & H.R. Siegismund 2008a. Three reciprocally monophyletic mtDNA lineages elucidate the taxonomic status of Grant's gazelles. *Conservation Genetics* 9: 593–601.
- Arctander, P., P. W. Kat, R. A. Aman & H.R. Siegismund 1996. Extreme genetic differences between populations of *Gazella granti*, Grant's gazelle, in Kenya. *Heredity* 76: 465–475.

Some general advice

You will probably get stuck at some point due to software not working on your files, R code not working etc. This happens all the time even for experienced researchers. Don't panic; try to dissect the problem by running a simpler command (in R, run lines one by one to see exactly where it goes wrong), visualize the input data to check that it looks as you expect or run the same command on an example data set that is almost always supplied with software packages. Googling the error message along with the software name is also a good way to find help. Or use your fellow group members to trouble shoot. Most of the time you will find that the cause of the problem was a tiny detail in the format of the data, you have not correctly changed some R code to fit your particular input, a typo or something similar. If you are unable to resolve the problem after trying all of the above, contact us and we can try to help you. Instead of being frustrated by setbacks keep in mind: this is what real data analysis is like, and you always learn something from them.