

Geovisualization Challenges of Seascape Genetics

Eoin Mac Aoidh, Jann Th. Martinsohn

Abstract—Population genetics are widely used for wildlife management and can also support environmental policy decisions. Landscape genetics, termed 'seascape genetics' for the marine environment, relates the population genetic structure of a species to its surrounding spatial environment. This allows us to link the genetic make-up of a population with influential environmental factors. Unfortunately, there prevails a communication gap between the genetic research community and regulators in the domain thus valuable information resulting from land- or seascape genetics is often lost in translation. In this article we look at the fisheries sector as a case study and propose geovisualizations as an effective means to improve the communication of information resulting from seascape genetic analysis to regulators and other stakeholders in the domain. Using this example we discuss the challenges of implementing such geovisualizations to better integrate genetic data into fisheries management, and more generally to effectively communicate scientific advice to stakeholders.

1 INTRODUCTION

Genetic analysis can be used as a tool for species identification [7] and also to distinguish between different populations of a species [10]. Both abilities are particularly useful for fisheries management including control and enforcement. By analyzing the genetic characteristics of fish of the same species taken from different geographical regions, distinct clusters of fish with specific genetic patterns are identifiable. This clustering of fish with similar characteristics can be exploited for origin assignment, an important aspect in fisheries control and enforcement since intentionally falsified declarations of geographic origin of fish and fish products frequently occurs. An array of analyzed specimens are used to generate a baseline, defining the locations associated with each genetically similar cluster. Following the identification of specific genetic characteristics in a specimen taken from a fisherman's catch, the specimen can be assigned as originating from a specific location with a given probability [6]. This technique was recently used to prosecute a fisherman who claimed his catch was from the Baltic Sea, where he was legally entitled to fish. Following a DNA analysis it was proved beyond reasonable doubt that the fish in the catch could not have been caught in the Baltic and had been caught in the North Sea, where the fisherman had engaged in illegal fishing. The fisherman was prosecuted accordingly [1]. As such, identifying and mapping the genetic structures in the ocean is particularly useful in the fight against Illegal Unreported and Unregulated (IUU) fishing, which has an estimated annual value of 10 - 20 billion euro [4].

It is well documented in the field of landscape genetics [8] that the characteristics of the landscape have an effect on the emerging genetic structure of the animals inhabiting the landscape. For example, a population of deer separated by an obstacle such as major motorway or large mountain are unlikely to mix for breeding purposes, thus genetic differences are likely to appear in the populations on either side of the obstacle. Landscape genetics combines information on genetic analysis with landscape features and spatial statistics. Seascape genetics follows the same rationale. Clusters of genetic similarity are usually separated by intrusive land barriers, strong ocean currents, distance, changes in salinity levels, chlorophyll concentration, depth, food sources, etc. There are many uses for genetic analysis, including control and enforcement, fraud detection, management schemes in support of sustainable fishing, selective breeding and monitoring of

endangered species all stand to benefit. Nevertheless, the widespread uptake of genetic technologies by fishery stakeholders other than scientists is minimal. There is a poor communication of the output of genetic research in a currency that can be used by fisheries managers and control authorities [11].

The FishPopTrace consortium [2] (of which the authors are members) has performed a genetic analysis on over 7,500 specimen of four commercially important species; cod, hake, herring, and sole, sampled across the Mediterranean into the Atlantic, Barents Sea, North Sea and Baltic Sea. With this information, a baseline for the purposes of origin assignment has been created. The genetic structure unveiled by these species and its relationship with other features of the seascape has the potential to make a significant difference in the course of decision making, for example, in support of sustainable fishing. However, this assumes an effective communication of the genetic structure to the decision makers. Therefore, in this abstract it is the geovisualization of clusters of genetic variability, and their relationship with influential environmental factors in the maritime domain which are of concern. By producing effective geovisualizations of the marine genetic structure and its influences, we wish to bridge the communication gap and to have such information considered in the decision making process.

2 VISUALIZATION DATA CONSIDERATIONS

The choice of visualization technique is influenced by a number of factors. We wish to raise awareness amongst stakeholders about genetic diversity. Furthermore the visualizations should serve as a decision support tool. Thus it must be accessible, interactive and user friendly, allowing the querying and highlighting of information matching a set of given criteria. To this effect we have developed a Web-based GIS tool based on the OpenGeo stack [3]. The critical aspect of the tool is in the graphical visualization approach employed. It must provide visual imagery that is quickly comprehensible to users of highly varied backgrounds. The interactive nature of the Web-application allows the user to highlight and query aspects matching user-specified criteria within the visualizations.

In addition to the audience centric and functional requirements of the visualizations, a significant limitation lies within the available data. The quantity and distribution of locations sampled for genetic baseline creation represent only a tiny fraction of the entire ocean coverage. Obtaining sample specimen requires cooperation from fishermen and/or specific research vessel trips, which is costly. There are considerable man hours required, often in harsh environments. As such, there are large swathes of ocean which are not covered by genetic population analysis. Consequently, despite the novel insight genetic analysis provides into the distribution of fish populations in our oceans, it is strongly constrained to the areas in which sampling and analysis have been performed. As such, inferences about the genetic characteristics of our target species in unsampled waters may be difficult and have a low certainty. To draw conclusions about the population structure

- *European Commission - Joint Research Centre,
Maritime Affairs Unit,
Italy
E-mail: eoin.mac-aoidh@jrc.ec.europa.eu,
jann.martinsohn@jrc.ec.europa.eu*

Manuscript received 31 March 2010; accepted 1 August 2010; posted online 24 October 2010; mailed on 16 October 2010.

For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

spanning the whole area under investigation, we rely on the baseline which has been created by carefully choosing appropriate sampling locations. To this effect, geovisualization also provides valuable support in designing and evaluating sampling strategies.

Having taken these data considerations into account, a number of challenges have been encountered in visualizing the relationship between multiple genetically diverse population clusters, and the influence of the surrounding seascape. These challenges are briefly outlined in the following section.

3 CHALLENGES AND PROPOSED SOLUTIONS

It is the aspect of the assignment probability of a specimen to a genetic cluster which is of immediate concern as a visualization output. For example, if a fish specimen is sampled by a control authority at a fishing port, given its genetic characteristics, we would ideally like to be able to express that there is a probability of 99% that it originated at location X, 1% at location Y and so on. Not only is there a need to show the probability of fish from each of the baseline locations of belonging to each cluster, there is also the need to show the inverse relationship. In other words, the probability that a fish from a given location has been incorrectly assigned to each sampling location in the baseline. The directionality and distance (both spatial and genetic) of other genetic clusters, and other components of the seascape such as temperature and salinity might also be of interest, hence there are multiple variables to be represented. There are many documented ways to represent multiple variables on a single map as described for instance in Slocum et al. [5]. However, a suitable representation technique showing the relationships we wish to highlight has not been identified in any of the previously documented approaches.

An initial solution proposed is to employ contour maps representing the probability of fish at all locations of belonging to a specific cluster. Although visually appealing, there is an inherent problem in the use of contour maps for our purpose, as they are designed to represent a continuous plane. The probability of a fish living on the land is 0, therefore there is a need for holes in the continuous probability surface. Simply clipping a continuous contour surface according to land intersection does not produce the correct visualization of probability. The isolines must be contoured around the land to achieve the correct visualization. A downside of this approach is that inter-cluster relationships cannot readily be represented by such a visualization as each cluster comprises a full coverage layer. The layering of multiple full coverages is difficult for the eye to comprehend.

Another challenge which we address is to show the mix of cluster characteristics present at each sampling location. An initial approach was to plot the ratio of cluster proportions location as a series of pie charts, however, directionality and distance are difficult to take into account in this context. Furthermore, legibility of the pie charts may be difficult with particularly small ratios. An interim solution which has been pursued is to represent the mixing between clusters as a mix of colors. For example if a sampling location is composed of 70% cluster 1, and 30% cluster two, the color representing the location is a mix of 70% one color and 30% of another. This solution is limited by the RGB and CMYK color models. Mixing of more than 3 or 4 clusters can not be shown. Furthermore, each of the colors are unique depending on the contributing statistics. With many different subtle color combinations on the map, the human eye has difficulty distinguishing between values, as documented in for the coloring of bivariate statistical maps [9]. In order to alleviate confusion, this temporary solution also includes labeled arrows between locations indicating direction and probability of assignment to an incorrect cluster.

An alternative solution, currently being pursued, which will be described in detail in a full paper is a type of distance-distorted star map, with a 'star' at each sampled and analyzed location. The 'star' represents the ratios of contribution from each genetic cluster for the species, while also illustrating the relative geographic distance between cluster centers and their direction of origin from the location of the star, which corresponds to where the samples analyzed were caught. The points of the 'star' represent each genetic cluster. Their distance apart represents genetic dissimilarity between clusters at the

sampling location (star centroid), while the direction of the geographic location of origin of the cluster is represented in the direction in which the arm points. An example of such a star at a sampling location to the north of Scotland is shown in Figure 1.

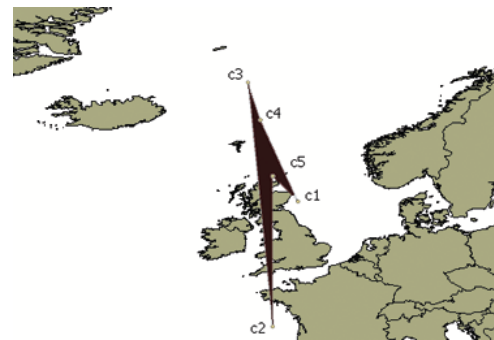


Fig. 1. Star map object illustrating the likely composition of the genetic structure of cod caught at a location to the north of Scotland. The star shows the quantity of contribution of genetic material from 5 clusters of genetically diverse cod populations. The points represent the direction in which the clusters are based. The length of the 'arms' illustrate the proportion of contribution from a given cluster

4 CONCLUSION

Genetics is a useful tool in many domains. We take fisheries management and control and enforcement as an example case study. A major barrier to its widespread uptake is the effective communication gap between scientists and decision makers. We wish to make the outputs of maritime genetic research more accessible to decision makers by providing tangible visualizations of the genetic structure of commercially important species. In this abstract we have outlined some of the important data considerations and challenges for the visualization of this genetic data and its relationship with the seascape. We will detail and evaluate the specific visualization approaches adopted in a full paper.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventy Framework Programme (FP7/2007-2013) under grant agreement no. KBBE-212399 (FishPopTrace).

REFERENCES

- [1] Danish directorate of fisheries, inspectorate of fisheries, technical university of denmark, 2006.
- [2] <http://fishpoptrace.jrc.ec.europa.eu/>, November 2010.
- [3] <http://opengeo.org/>, November 2010.
- [4] D. Agnew, J. Pearce, G. Pramod, T. Peatman, R. Watson, J. Beddington, and T. Pitcher. Estimating the worldwide extent of illegal fishing. *PLoS One*, 4(2):4570, 2009.
- [5] F. Kessler, R. McMaster, H. Howard, and T. Slocum. Thematic cartography and geographic visualization. 2007.
- [6] E. Nielsen, M. Hansen, C. Schmidt, D. Meldrup, and P. GrønkJaer. Fisheries: population of origin of Atlantic cod. *Nature*, 413:272, 2001.
- [7] F. Pereira, J. Carneiro, and A. Amorim. Identification of species with DNA-based technology: current progress and challenges. *Recent Patents on DNA & #38; Gene Sequences*, 2(3):187-200, 2008.
- [8] A. Storfer, M. Murphy, J. Evans, C. Goldberg, S. Robinson, S. Spear, R. Dezzani, E. Delmelle, L. Vierling, and L. Waits. Putting the 'landscape' in landscape genetics. *Heredity*, 98(3):128-142, 2006.
- [9] B. E. Trumbo. A theory for coloring bivariate statistical maps. *The American Statistician*, 35(4):pp. 220-226, 1981.
- [10] R. Waples and O. Gaggiotti. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology*, 15(6):1419-1439, 2006.
- [11] R. Waples, A. Punt, and J. Cope. Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries*, 9(4):423-449, 2008.