POPULATION GENOMICS

Evolution of cod supergenes

Population genomic and phylogenomic analyses of Atlantic cod provide new insights into the origin and maintenance of supergenes and highlight the role of recombination and structural variants.

Anna Tigano

The maintenance of different complex phenotypes within a population has long puzzled ecologists and evolutionary biologists alike. In recent years, genomics has enabled us to investigate the genetic basis and architecture of these polymorphisms, which is fundamental to understanding how they evolve. A rapidly increasing number of genomic studies have shown that complex polymorphisms are frequently associated with 'supergenes', genetic architectures able to maintain a tight association between genetic elements controlling different traits. Writing in *Nature Ecology & Evolution*, Matschiner et al.[1] provide new insights into the origin and maintenance of supergenes in Atlantic cod that are associated with migratory and stationary ecotypes and different adaptations to salinity and temperature across their range.

The power of supergenes lies in their ability to prevent sets of co-evolved loci from getting shuffled by gene flow and recombination, which is exactly what structural variants are great at doing (Fig. 1a). Indeed, supergenes are generally identified as a block-like pattern of differentiation and linkage disequilibrium, which abruptly plummet around the genomic region associated with the polymorphism (Fig. 1a). This pattern is consistent with the recombination suppression effect of structural variants, such as inversions or insertion/deletions, that are overwhelmingly found associated with complex polymorphisms and supergenes (Fig. 1a)[2]. This is also the case for the four supergenes in Atlantic cod investigated by Matschiner et al., which are all associated with inversions[3,4].

Matschiner and colleagues go well beyond the detection of inversions and look in depth into their recombination suppression role in the origin and maintenance of these supergenes. For example, they report that patterns of linkage disequilibrium and the breakpoints of these inversions generally match very closely, but not always, showing that the recombination suppression effect of inversions can
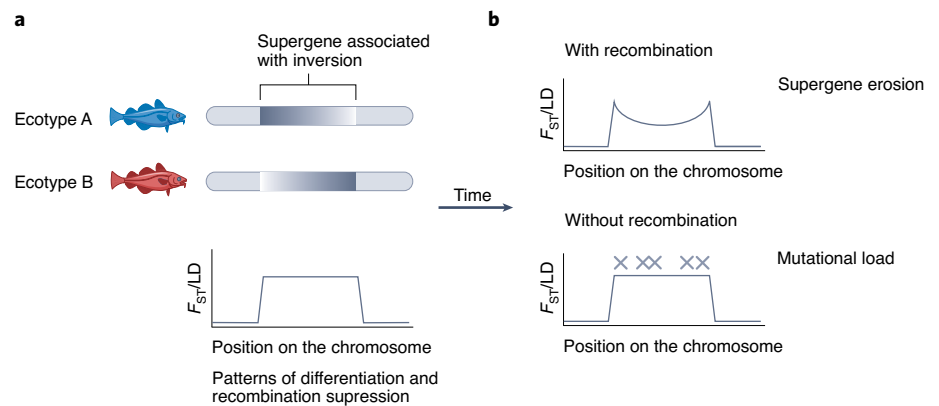

Fig. 1 | Evolution of supergenes. Supergenes are generally identified by block-like patterns of differentiation and linkage disequilibrium (LD), more often than not associated with structural variants, which is the case for the four supergenes associated with inversions in the Atlantic cod studied by Matschiner et al. **a**, A supergene associated with two ecotypes results in increased population differentiation (F_{ST}) between the two ecotypes and linkage disequilibrium within each of the two haplotypes. **b**, With time, supergenes are expected to erode and experience a reduction in divergence, if recombination occurs between the two alternative haplotypes (top), or to accumulate deleterious mutations (depicted by crosses), if recombination does not occur (bottom). The four supergenes in the Atlantic cod show no signs of either, suggesting that low levels of recombination keep the supergenes 'healthy'.

extend beyond their breakpoints. Though recombination breaks the association of co-evolved loci within supergenes, eventually causing their erosion (Fig. 1b), it also covers the important function of removing deleterious mutations, which can accumulate unchecked when recombination is suppressed (Fig. 1b). Matschiner et al. do not find evidence for mutation load within inversions nor for supergene erosion due to recombination, but their results support allele exchange (gene flux) between haplotypes with opposite arrangements via gene conversion and double crossovers in two of the four supergenes. How, then, can all four supergenes look 'healthy', that is, with no signs of erosion nor mutation load? The authors exclude the hypothesis of a recent supergene origin by phylogenomic-based analyses that estimate divergence times between 0.40–1.66 million years ago. They suggest that even very low levels of gene flux could prevent the

erosion of supergenes[5], but it is also possible that inherent difficulties in detecting gene conversion or double crossovers, especially if occurring at low rates, could lead to an underestimation of gene flux in supergenes. Furthermore, although the potential role of demographic factors in the maintenance of supergenes is not discussed and remains to be tested, large effective population sizes could provide plenty of recombination opportunities among individuals with the same arrangements. Atlantic cod were historically very abundant, until they were overfished during the past century.

Matschiner et al. also perform a set of elegant analyses to investigate the origin of supergenes in cod. The authors first reconstructed phylogenetic relationships across four cod species, and used this time-calibrated phylogeny to test whether supergenes introgressed from other species, as other studies have shown[6]. Supergenes in Atlantic cod do not appear

to have been introduced via gene flow from a closely related species, suggesting that the evolution of these supergenes occurred after speciation. Indeed, the authors detected and dated signatures of a bottleneck in the derived arrangement of all supergenes, coinciding or following the supergene origin, which lends support for a single origin for each of the four supergenes. What's interesting here is that the effective population size inferred from these supergene haplotypes was seemingly restored to pre-bottleneck levels pretty quickly, suggesting, once again, that demographic factors, such as the large historical population size of Atlantic cod, may be an underappreciated factor in the evolution and maintenance of supergenes.

Moreover, Matschiner and colleagues are able to determine what arrangements are ancestral or derived for each supergene. This piece of information proved useful to investigate the evolution of supergenes from a molecular point of view, and it could help understand what environmental changes promoted the evolution of these polymorphisms in the past, and even predict the effect of projected environmental changes on the frequency of

these different ecotypes in the future. For example, we can ask how the frequency of alternative arrangements on LG 2, 7 and 12, which are associated with different salinity levels and temperatures, will change with climate change.

Supergenes are now relatively easy to identify, however, as Matschiner et al. show, their genomic structure can limit their study. First, because of strong linkage and recombination suppression within supergenes, disentangling the effect of different variants on different genes or regulatory elements on the various traits contributing to the phenotypic polymorphisms is next to impossible. Therefore, causation is hard to demonstrate. Second, supergenes evolve somewhat independently from the rest of the genome and from each other, meaning that selection, drift, gene flow, recombination and even mutation rates can be dramatically different within and outside supergenes[7]. Despite the authors' best effort to date supergenes, and the comprehensive phylogenomic analyses they perform, the divergence dates for supergene haplotypes should be taken with a grain of salt, at least until we better understand the effect of structural variants

on molecular evolution and of how to take those into account for phylogenomic reconstructions[7,8]. Nonetheless, Matschiner et al. show the importance of structural variants and recombination in the evolution of phenotypic divergence, adaptation and incipient speciation. ❏

Anna Tigano 🆔 ✉

Biology Department, University of British Columbia, Okanagan, Kelowna, British Columbia, Canada.
✉*e-mail:* anna.tigano@ubc.ca

References

1. Matschiner, M., Barth, J. M. I., Torresen, O. K. & Star, B. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-022-01661-x (2022).
2. Gutiérrez-Valencia, J., Hughes, P. W., Berdan, E. L. & Slotte, T. *Genome Biol. Evol.* **13**, evab057 (2021).
3. Kirubakaran, T. G. et al. *Mol. Ecol.* **25**, 2130–2143 (2016).
4. Kirubakaran, T. G. et al. *G3* **10**, 2903–2910 (2020).
5. Yan, Z. et al. *Nat. Ecol. Evol* **4**, 240–249 (2020).
6. Jay, P. et al. *Curr. Biol.* **28**, 1839–1845.e3 (2018).
7. Faria, R., Johannesson, K., Butlin, R. K. & Westram, A. M. *Trends Ecol. Evol.* **34**, 239–248 (2019).
8. Mérot, C., Oomen, R. A., Tigano, A. & Wellenreuther, M. *Trends Ecol. Evol.* **35**, 561–572 (2020).