

# Inversions or Introgression?

## The origin of high-divergence genomic regions in Atlantic cod

Matschiner M<sup>1</sup>, Star B<sup>1</sup>, Berg PR<sup>1</sup>, Tørresen OK<sup>1</sup>, Nederbragt AJ<sup>1</sup>, Hansen MHS<sup>1</sup>, Skage M<sup>1</sup>, Kjesbu OS<sup>1,2</sup>, Stenseth NC<sup>1</sup>, Jakobsen KS<sup>1</sup>, Jentoft S<sup>1</sup>

<sup>1</sup>Centre for Ecological and Evolutionary Synthesis, Department of Bioscience, University of Oslo, Oslo, Norway

<sup>2</sup>Institute of Marine Research, Box 1870 Nordnes, N-5817 Bergen, Norway

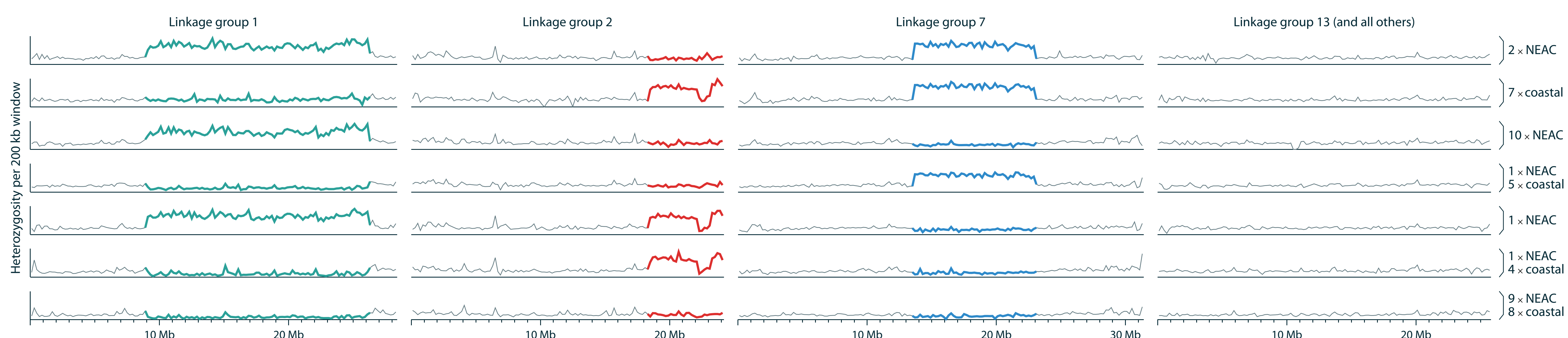
email: michaelmatschiner@mac.com

**Hi.** Thanks for your attention. This is an attempt to use the collective brain power of ESEB 2015 to tackle the question whether inversions or introgression are driving population divergence in Atlantic cod (*Gadus morhua*). Populations of

Atlantic cod in Norwegian waters differ in a variety of phenotypic traits, including alternate life history strategy and migration behaviour. In contrast to local cod populations which reside near the coast all-year round, the North-East Arctic Cod (NEAC) is a highly migratory population, feeding most of the year in the open water of the Barents Sea, but spawning along the Norwegian coast. Besides long-distance migration between the Barents Sea and the spawning area, the North-East Arctic Cod is further characterized by vertical migration and the use of deeper and more offshore habitats, as well as the crossing of thermal fronts. To determine the genetic basis of this phenotypic variation, we have sequenced 24 individuals of both the NEAC and coastal population to ~9X coverage, using Illumina HiSeq technology. What we find is surprising.

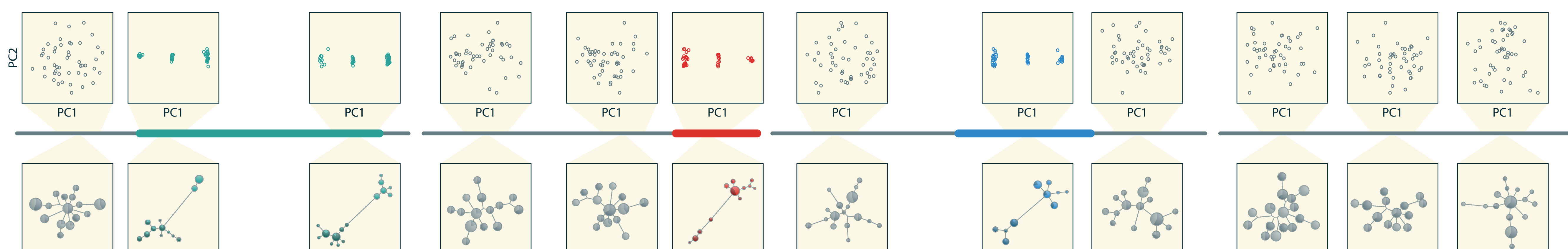
### 1) High heterozygosity regions are found in three linkage groups.

Paired-end reads of all individuals were mapped against the Atlantic cod genome reference gadMor2 (Tørresen et al, in prep.), and variant calling resulted in a total of 13.6 million SNPs. We find that heterozygosities of SNPs differ markedly across the genome, and between individuals. In regions of up to 17 million bases (Mb) on linkage groups 1, 2, and 7, the heterozygosity is highly elevated in some individuals, indicating the presence of pairs of long divergent alleles in each of these linkage groups. Notably, the pattern differs between the NEAC and coastal populations, as e.g. high heterozygosity in linkage group 1 is only found in the NEAC population.



### 2) Genetic variation is highly structured in high-heterozygosity regions.

Within high-heterozygosity regions, principal component analyses of 3 Mb alignment windows show three genetic clusters, representing two major alleles that occur homozygously and heterozygously. Representative haplotype genealogies based on sets of 30 consecutive SNPs are shown below. Edge lengths indicate number of transversion needed to connect haplotypes, and only haplotypes occurring more than once are shown.



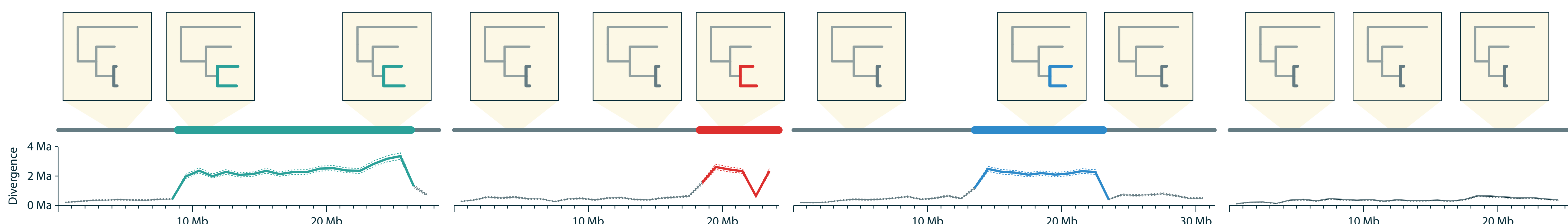
### 3) Boundaries of high-divergence regions do not seem to coincide with inversion breakpoints.

If high-divergence regions were caused by inversions carried by part of the population, we would expect to find long-distance links between read pairs where one of the two reads maps near the beginning and the other near the end of the region. In fact, we find such links in all three linkage groups, but there is no correspondance with the beginning and end of the divergent alleles, and not with the individuals carrying them.



### 4) Divergent alleles of Atlantic cod are closer to each other than to the sister species, but 2-3 million years old.

If high-divergence regions were caused by introgression, we would expect one of the two alleles in each region to be closer to homologs from related species than to the other allele. We produced genome-wide alignments for two Atlantic cod individuals carrying the alternative alleles and nine newly-sequenced closely related species (including the sister species of Atlantic cod). Phylogenies of 1 Mb alignment windows show that the Atlantic cod individuals cluster monophyletically, inside and outside of high-divergence regions (only two out of nine outgroups shown in Figures). Time-calibration with BEAST2 indicates allelic divergence 2-3 million years ago.



**So,** what do you think? Is this a sign of inversions or introgression? Cast your vote here:

- Clearly inversions \_\_\_\_\_
- No way. Introgression \_\_\_\_\_
- Something else \_\_\_\_\_

