Supplementary VIa: Context dependent prediction in DNA sequence using neural networks. Fourier, yeast genome, LSTM50.

Christian Grønbæk, Yuhu Liang, Desmond Elliott, and Anders Krogh

April 16, 2022

This file contains two sets of plots based on the Fourier analysis showing the L2-norm of the Fourier coefficients in a running window. This first set covers the frequency range from 20 to 4500 using a window length of 100 (and a step size of 10), while the second covers the frequencies from 4000 to 14000 and used a window length of 500 (and step size of 10). Note that, since the window length is five times larger in the second set than in the first, the scale of the amplitudes (y-axis) is five times higher in the second set than in the first.

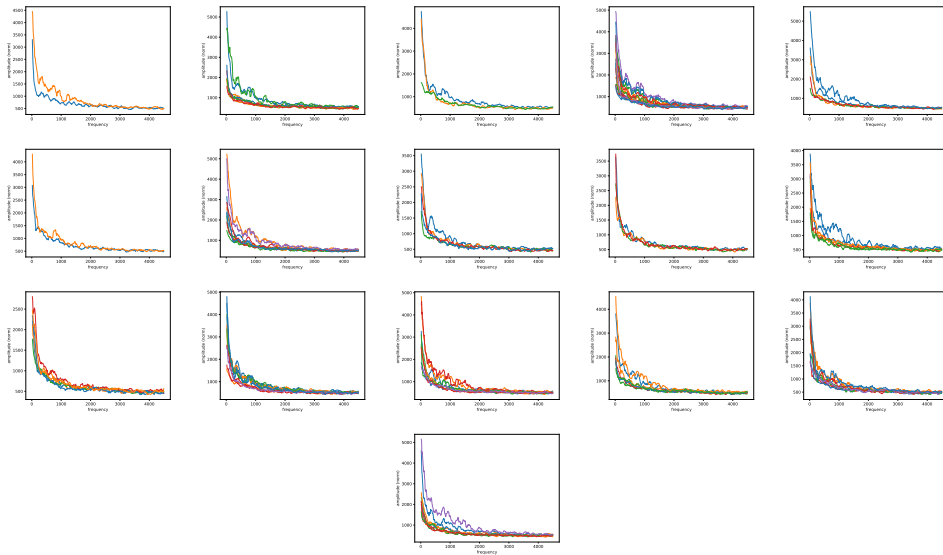# Fourier plots, LSTM50 reference base probabilities, frequency range 20 to 4500

Figure 1: Yeast LSTM50. Fouriers on reference-base probability, low frequency range. Each plot covers one chromosome, listed in increasing order (chr1 to chr16). The genome string is divided in adjacent segments of 100Kb (per chromosome); each plot shows the results for all segments in the chromosome (with ratio of qualified positions > 0.9).

Fourier plots, LSTM50 reference base probabilities, frequency range 4000 to 14000
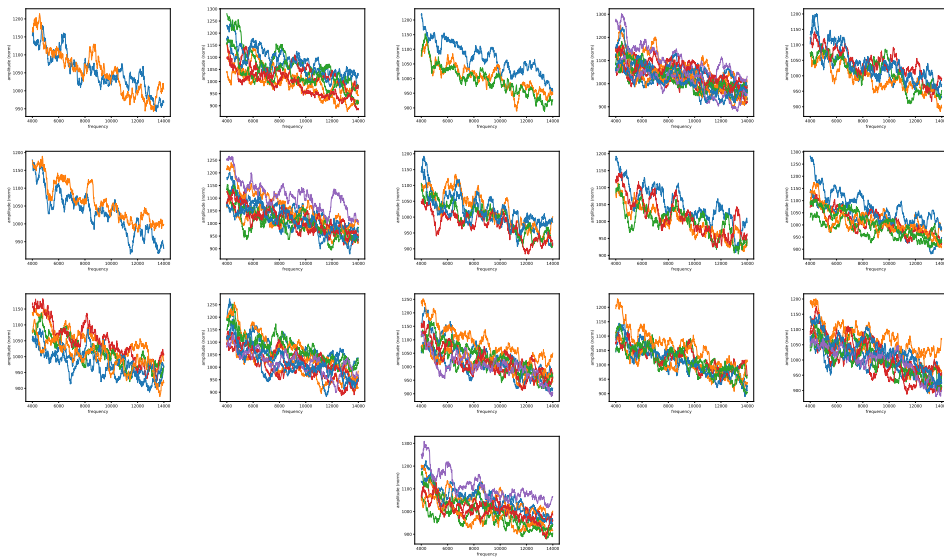

Figure 2: Yeast LSTM50. Fouriers on reference-base probability, higher frequency range. Each plot covers one chromosome, listed in increasing order (chr1 to chr16). The genome string is divided in adjacent segments of 100Kb (per chromosome); each plot shows the results for all segments in the chromosome (with ratio of qualified positions > 0.9).