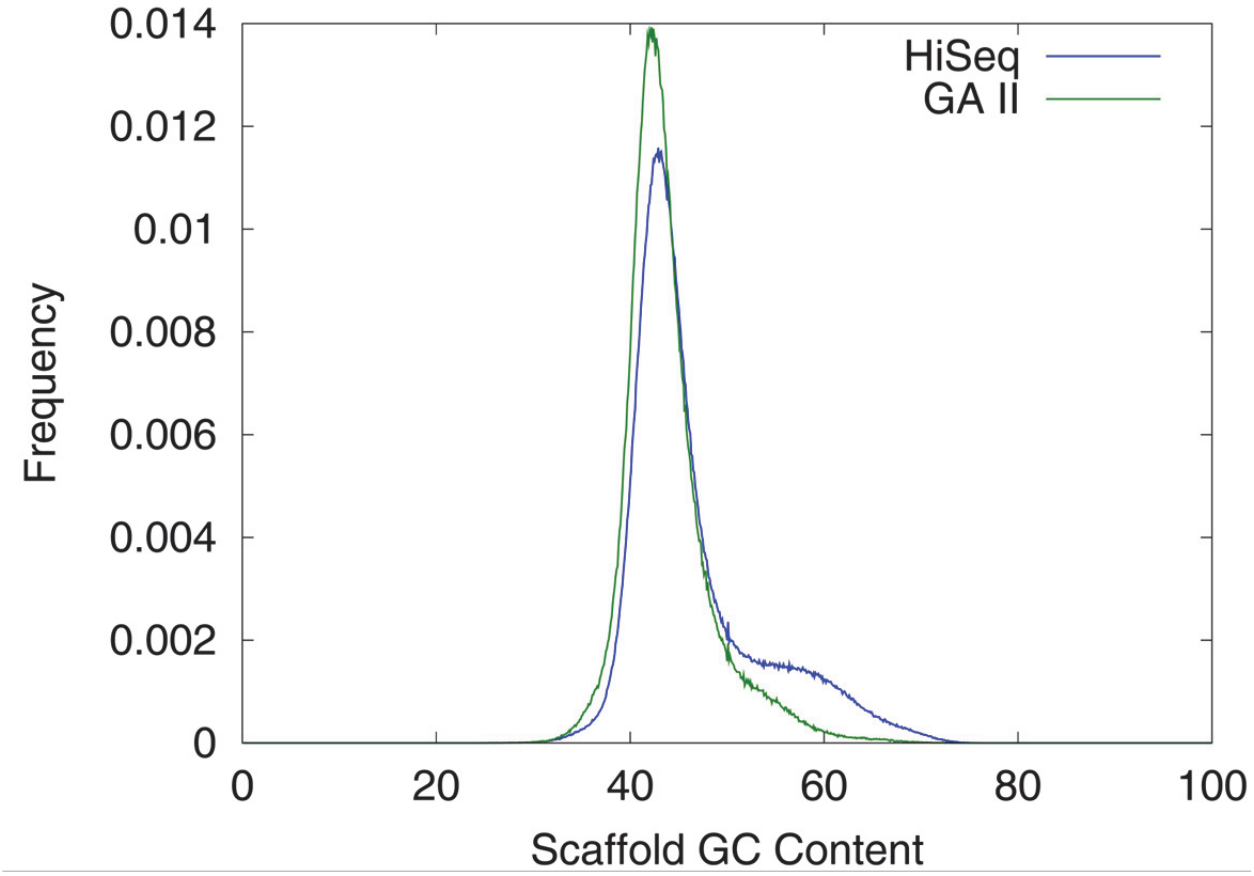

Figure S1. A comparison of the frequency of scaffolds according to variation in % GC content between samples sequenced on HiSeq versus GA II platforms. Distributions are broadly overlapping. The long right-tail from the HiSeq samples is caused by a disproportionate number of algae samples sequenced on that platform. These algae exhibited especially rich GC transcripts, which is consistent with the results of published whole genome sequences of green algae[40,41].