



**Figure S2.** The figures show how G-C content affects insertion index in different genomes in this study. Each dot represents a gene and the green line shows the GAM curve when the formula is  $y \sim s(x)$ . The light shade shows the 95% confidence interval.