

Appendix S2. Supplemental figures for “nQuack: An R package for predicting ploidal level from sequence data using site-based heterozygosity”

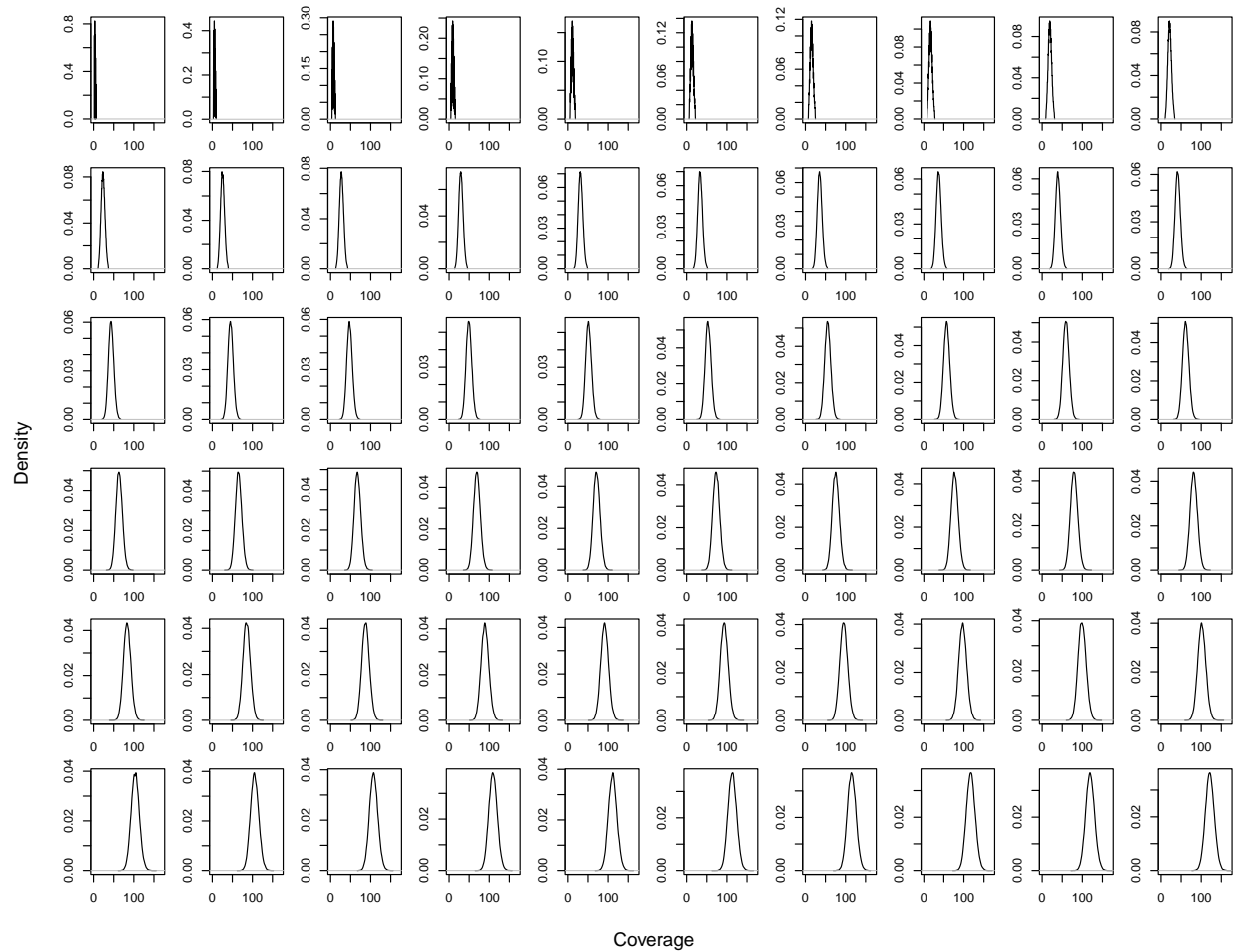


Figure S1. Distribution of sequence coverage of the realistic simulation approach.

Figure S2-S55. — Model comparisons for simulated data sets when considering all ploidal levels (Figures S2–S19) or only a subset of ploidal levels: (1) only diploids, triploids, and tetraploids (Figures S20–S37), or (2) only diploid, tetraploids, and hexaploids (Figures S38–S55). BIC score difference between the best and second-best models for simulated samples across different numbers of sites for 11 different coverage amounts (5x, 10x, 20x, 30x, ...). The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. A larger BIC difference between the best and second-best models should indicate model confidence; model confidence in incorrect mixtures can be observed in these plots. These plots can be used to guide users' interpretation of these models and determine whether these models will apply to their system.

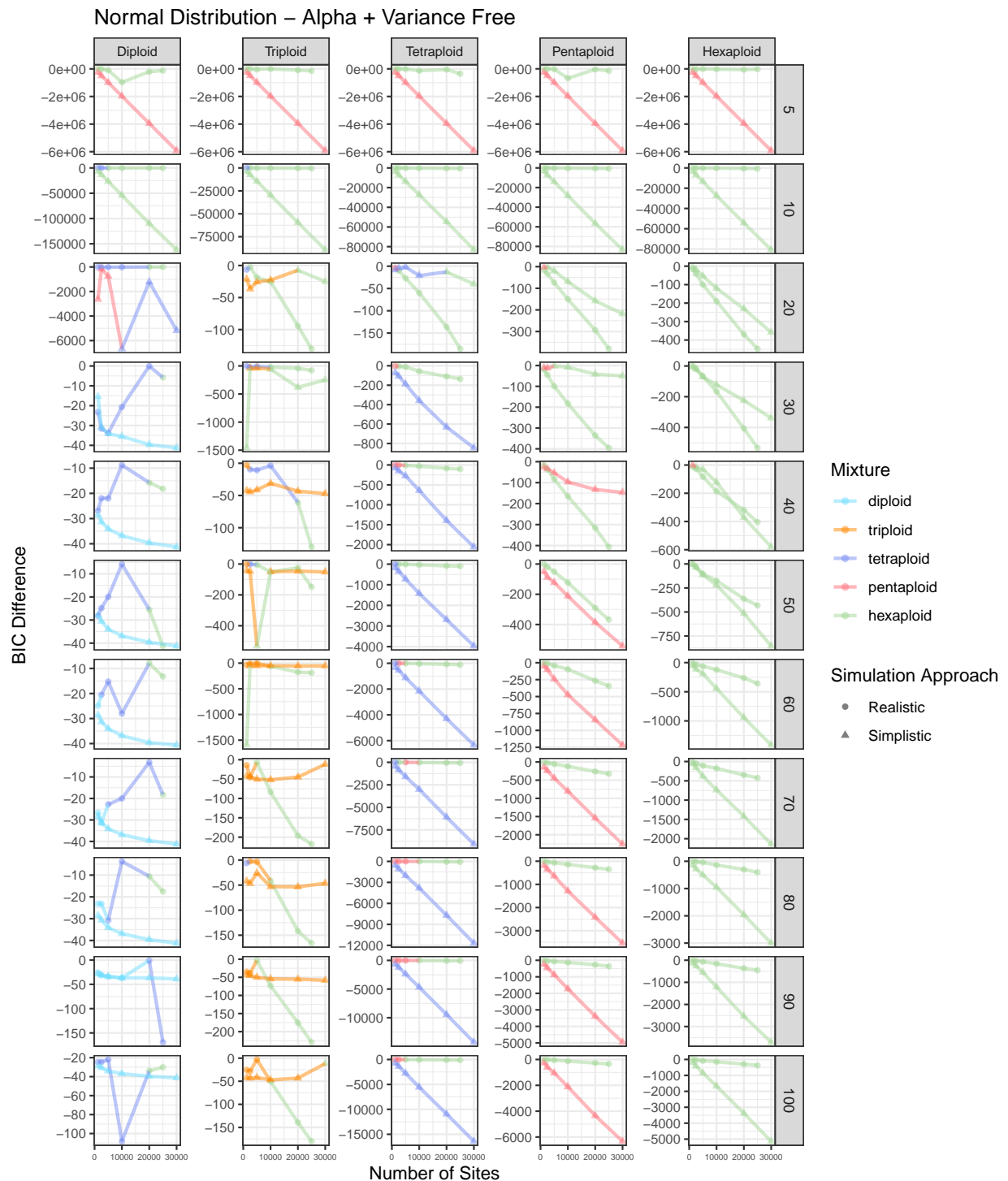


Figure S2. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha and variance free.

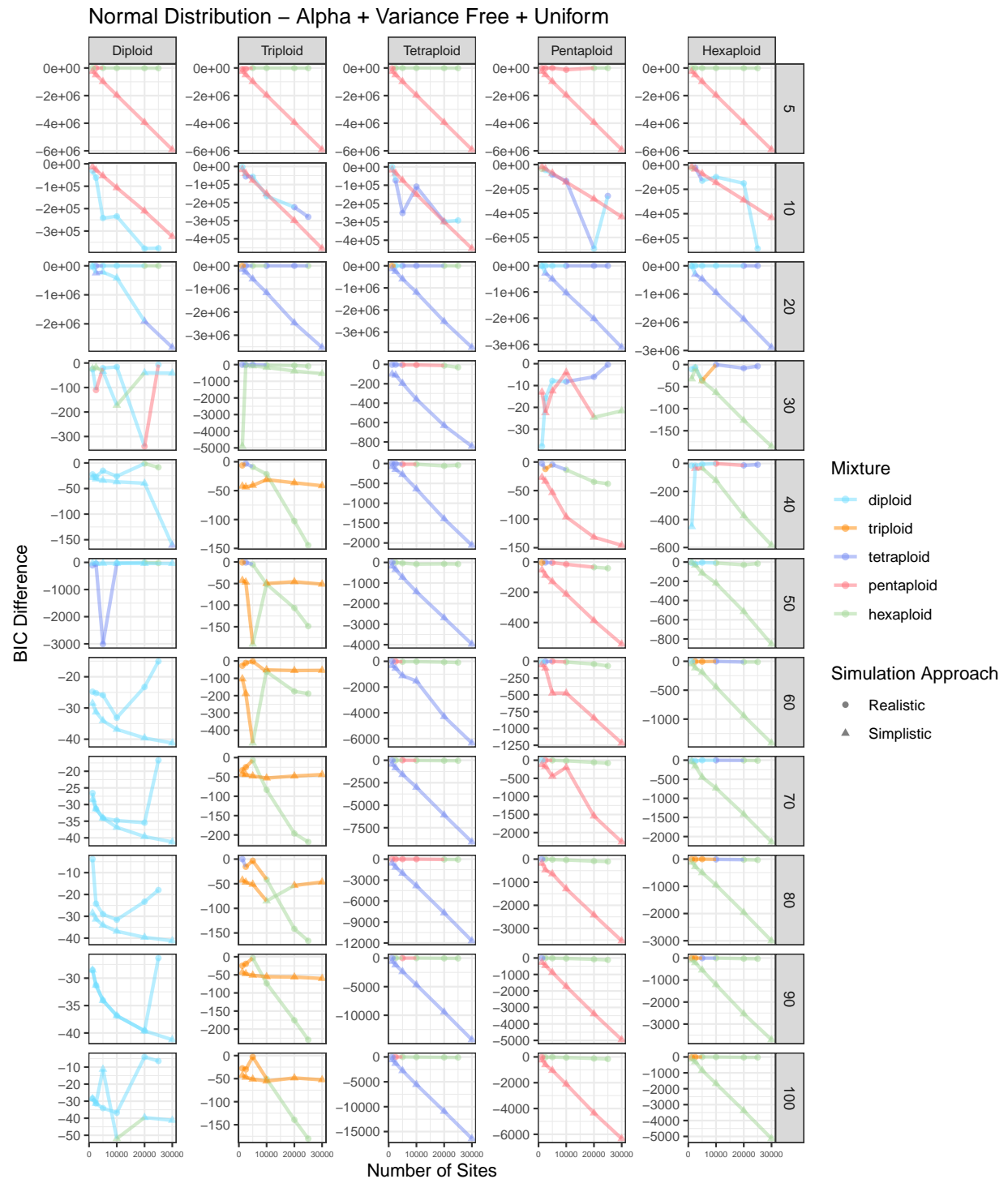


Figure S3. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha and variance free with a uniform mixture.

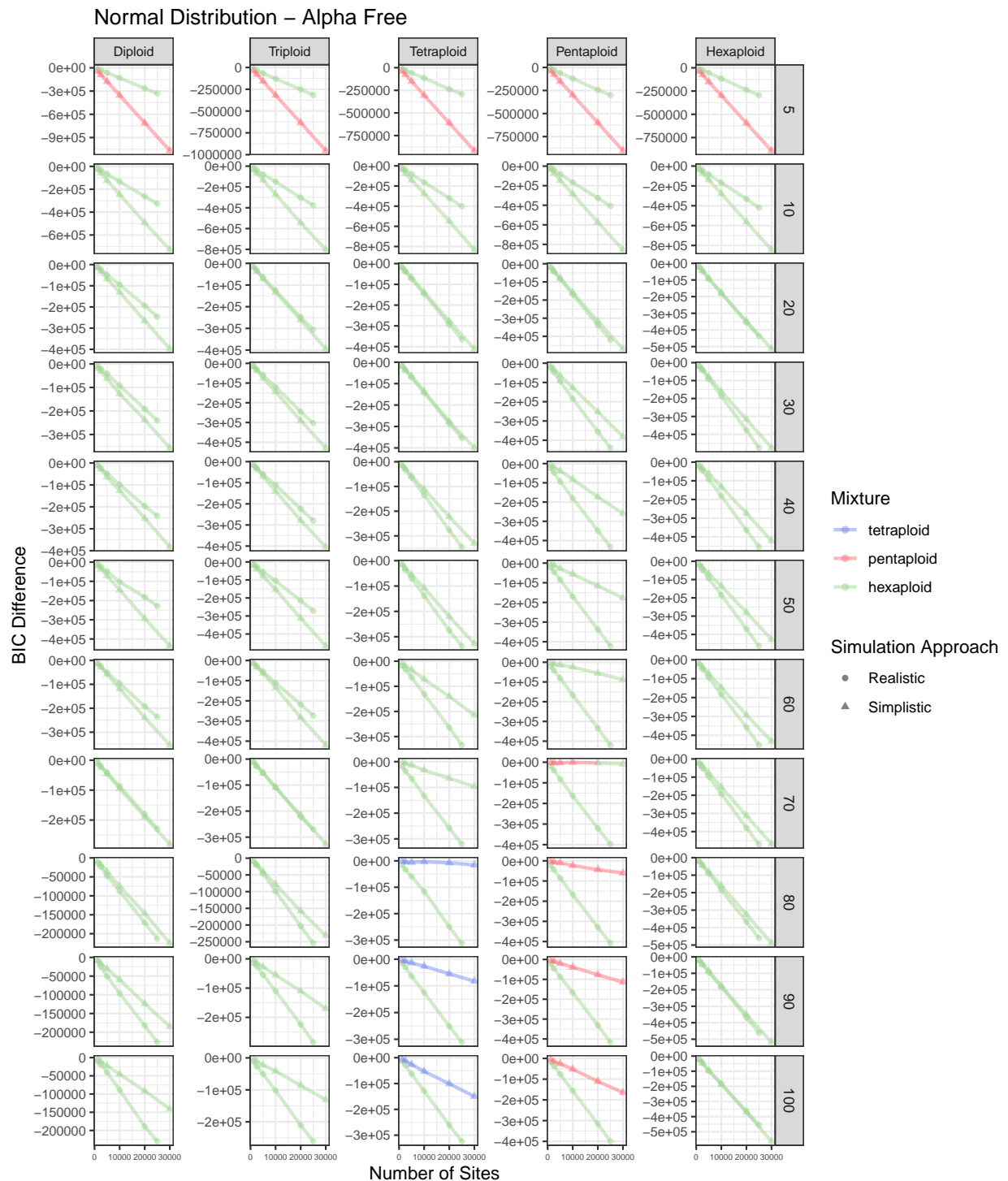


Figure S4. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha free.

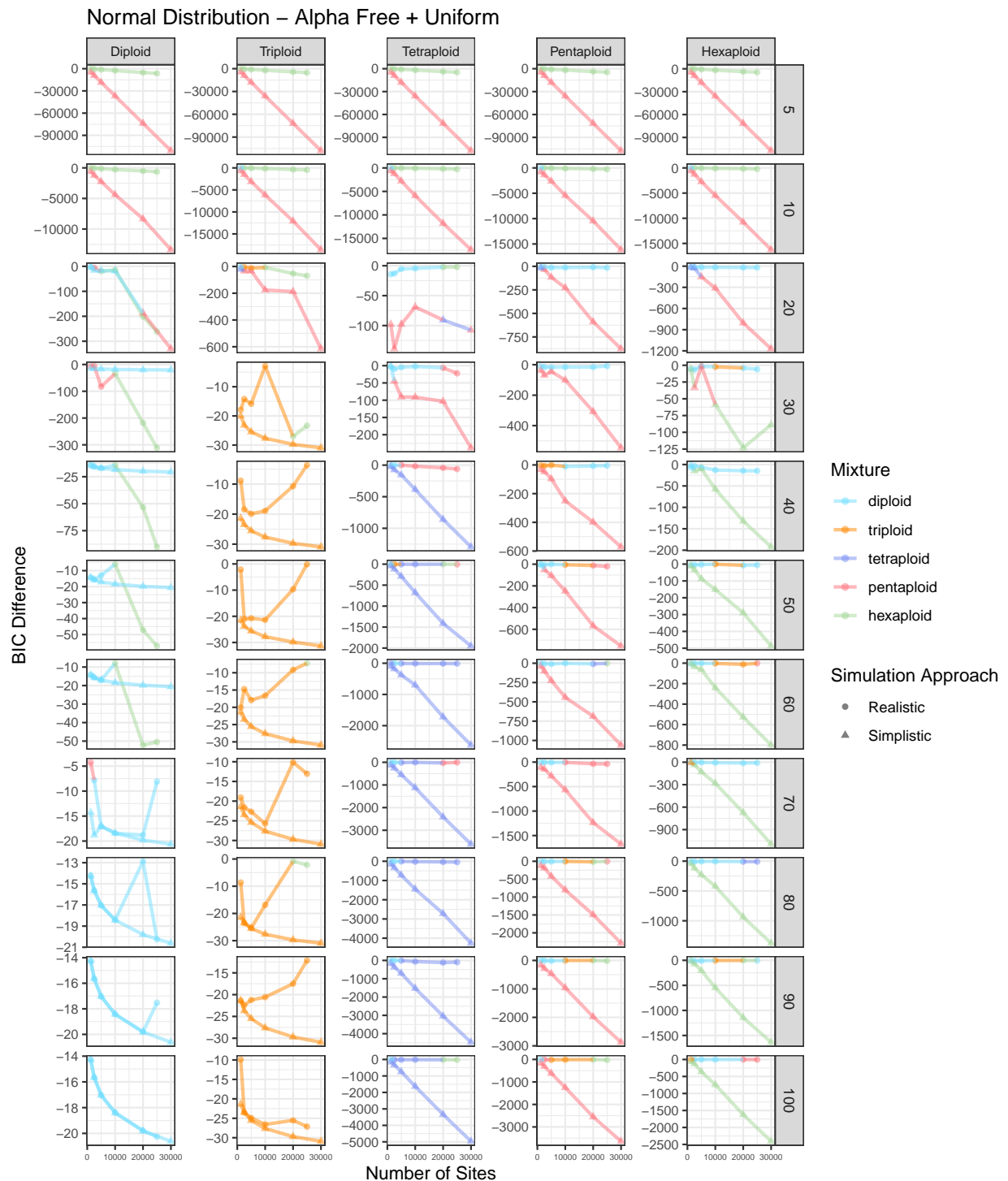


Figure S5. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha free with a uniform mixture.

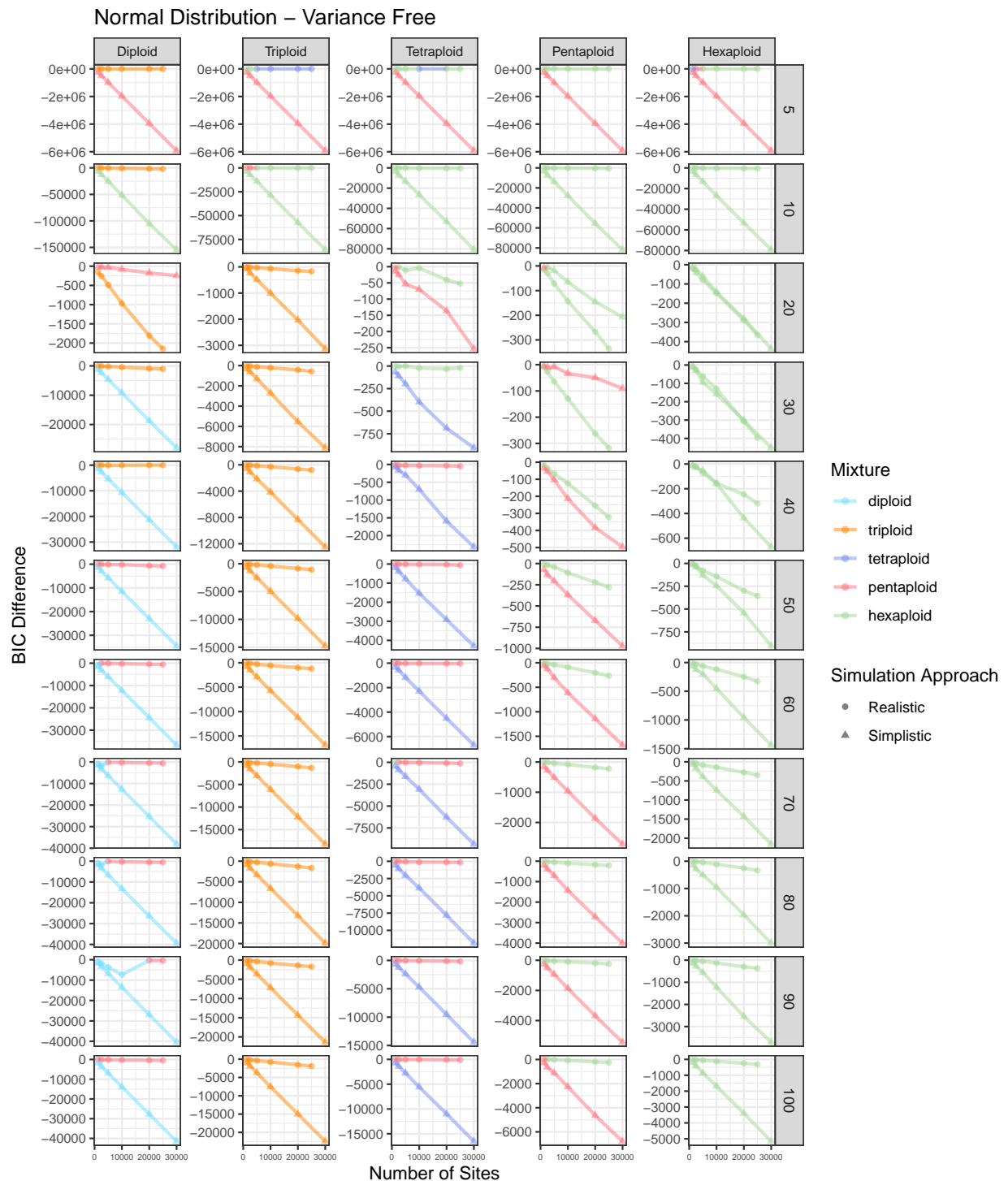


Figure S6. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with variance free.

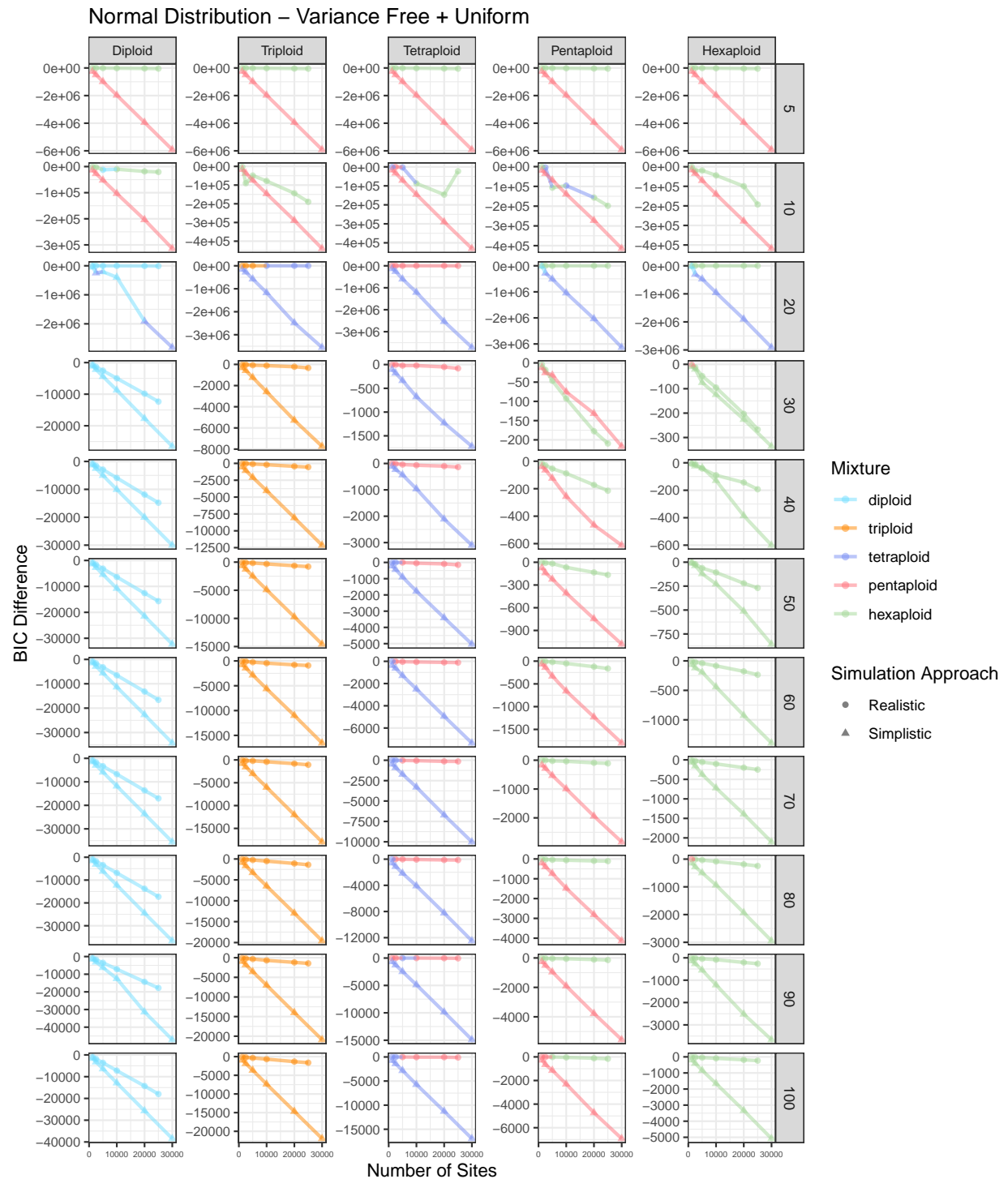


Figure S7. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with variance free with a uniform mixture.



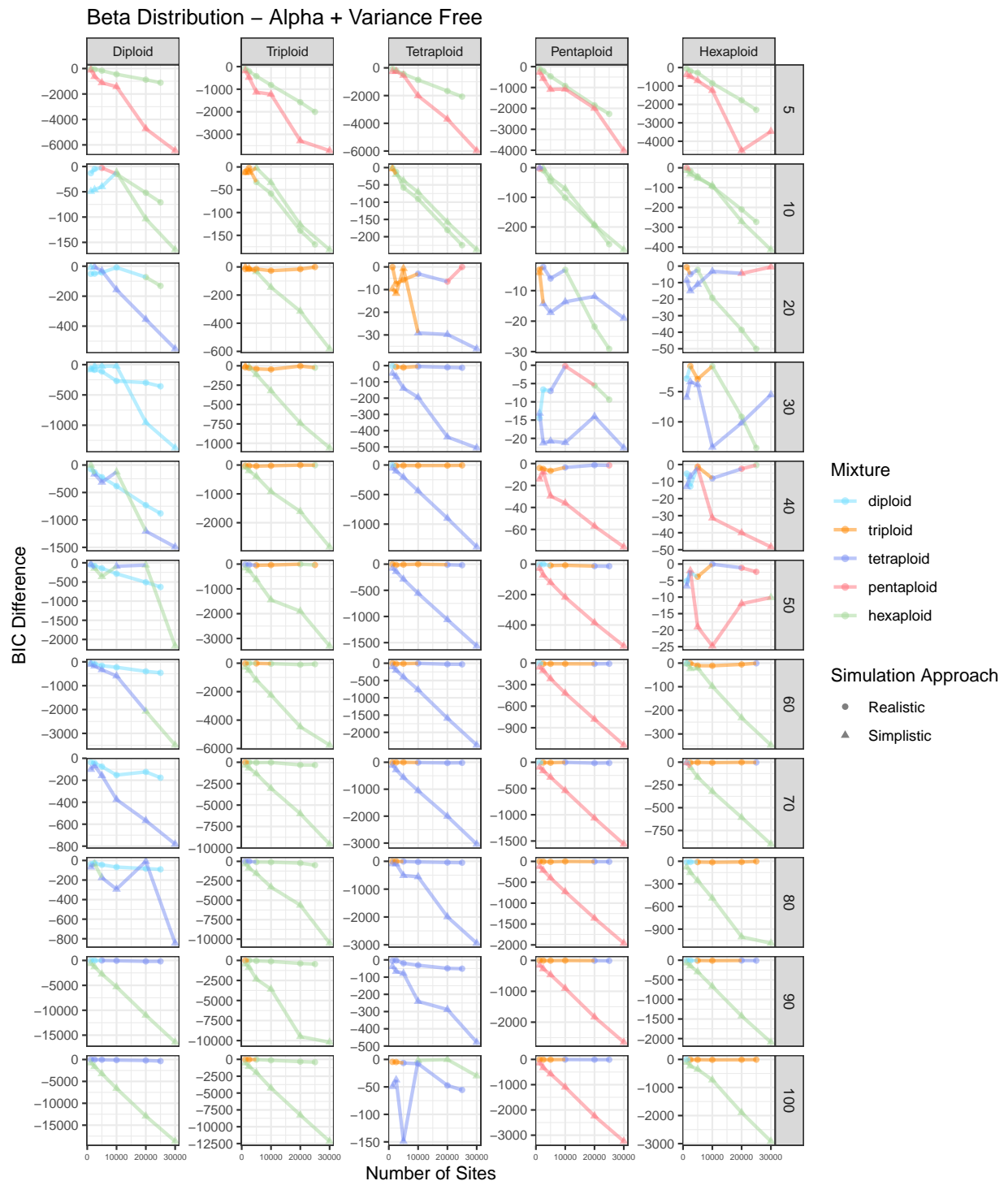


Figure S8. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha and variance free.

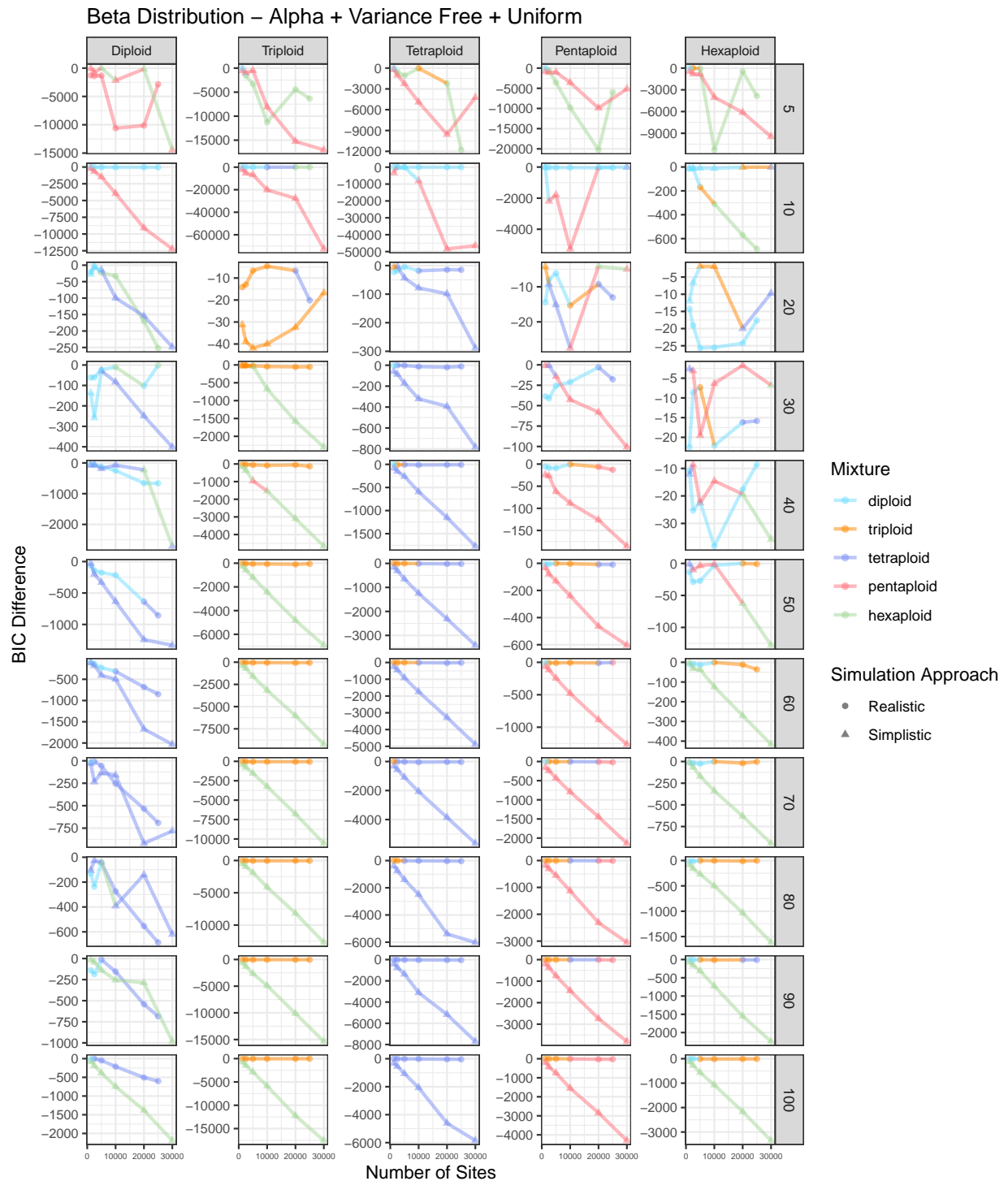


Figure S9. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha and variance free with a uniform mixture.

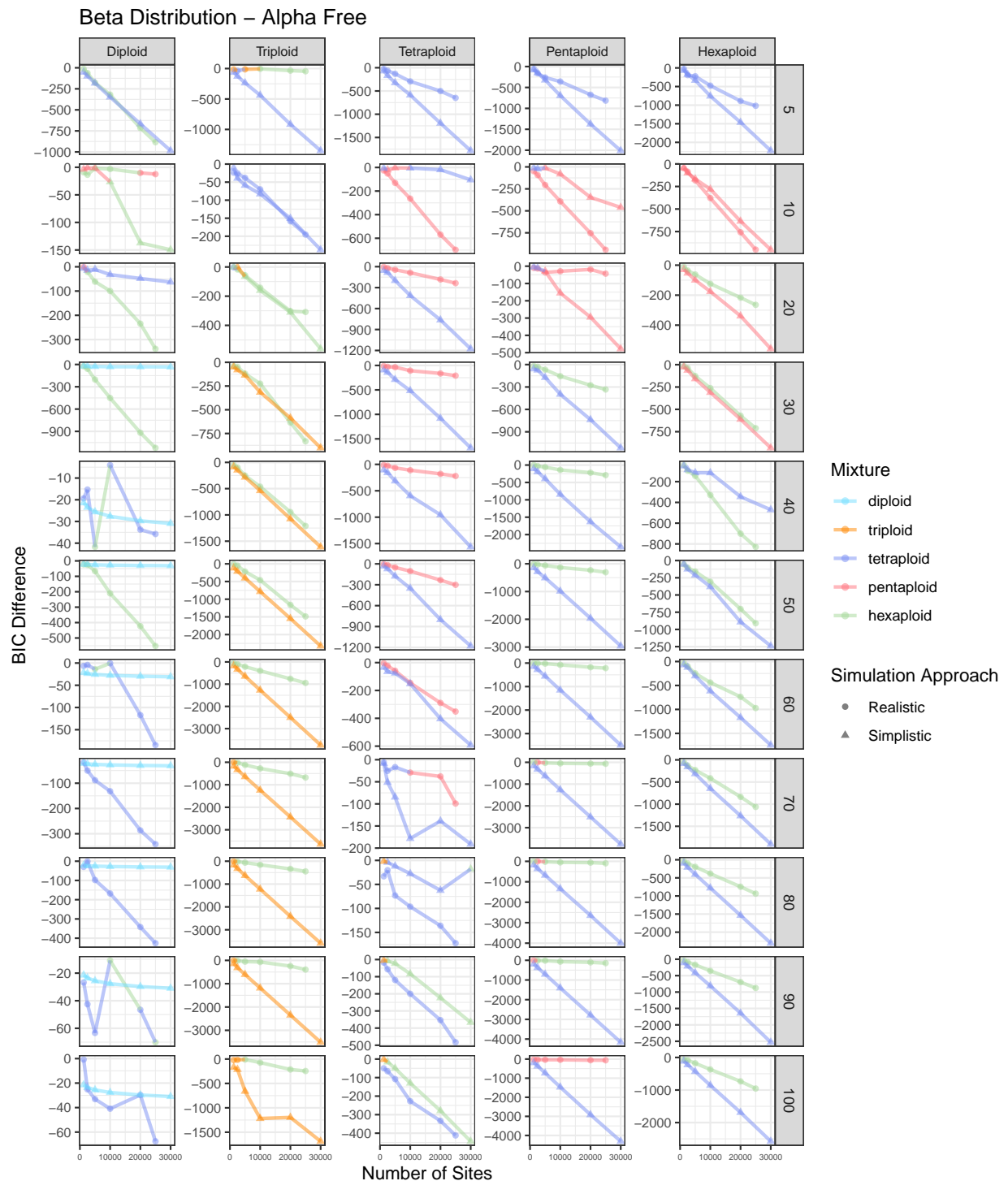


Figure S10. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha free.

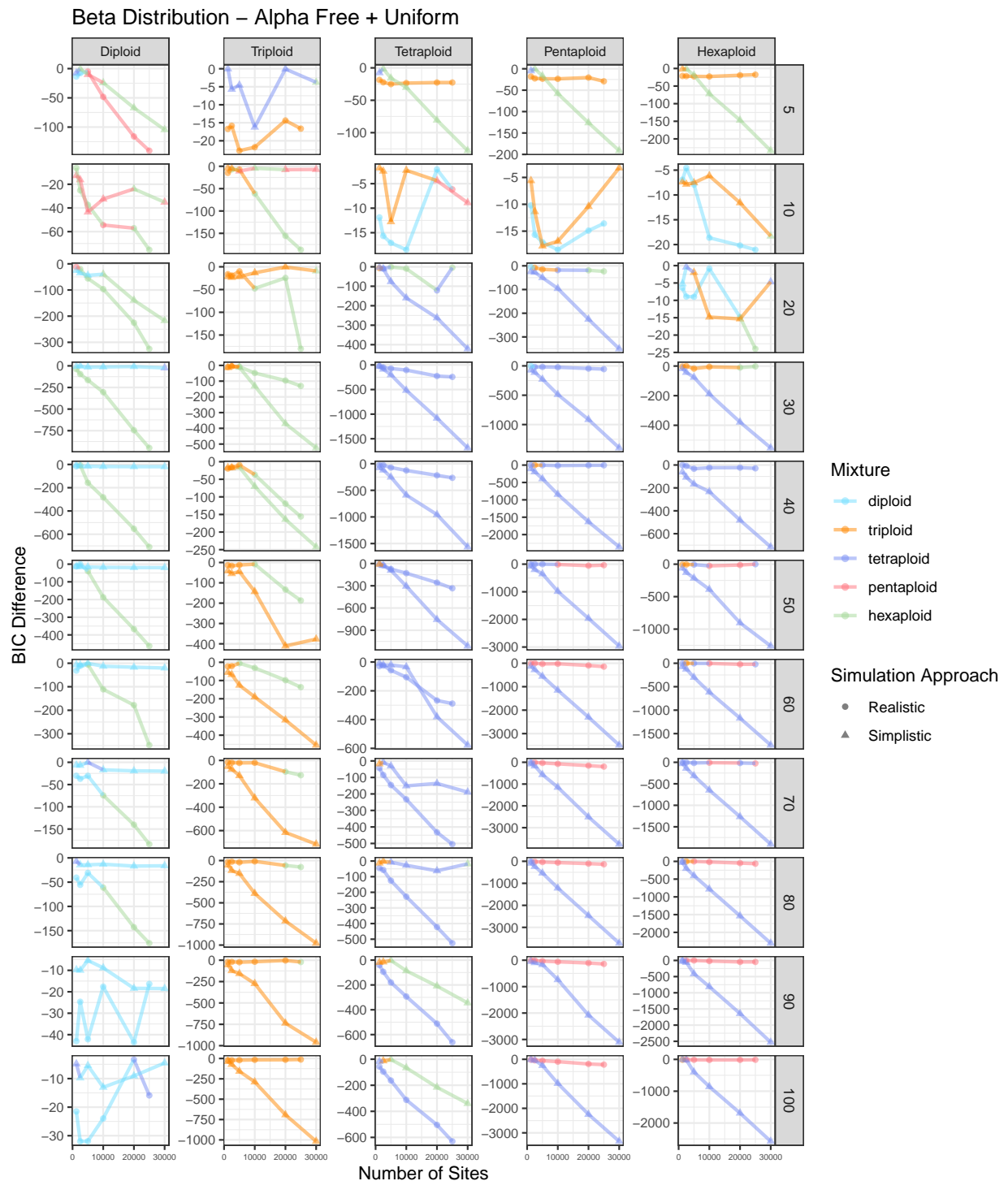


Figure S11. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha free with a uniform mixture.

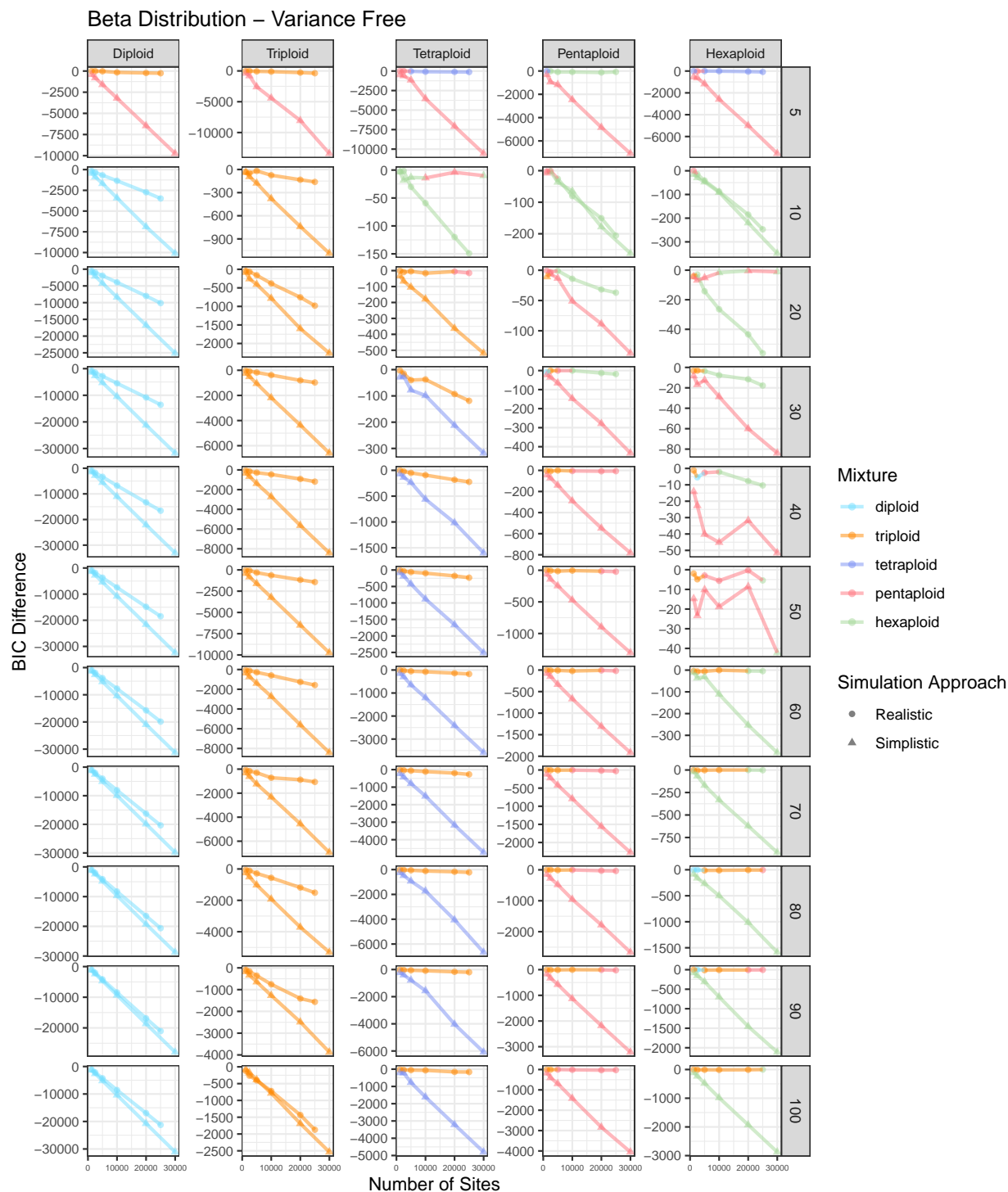


Figure S12. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with variance free.

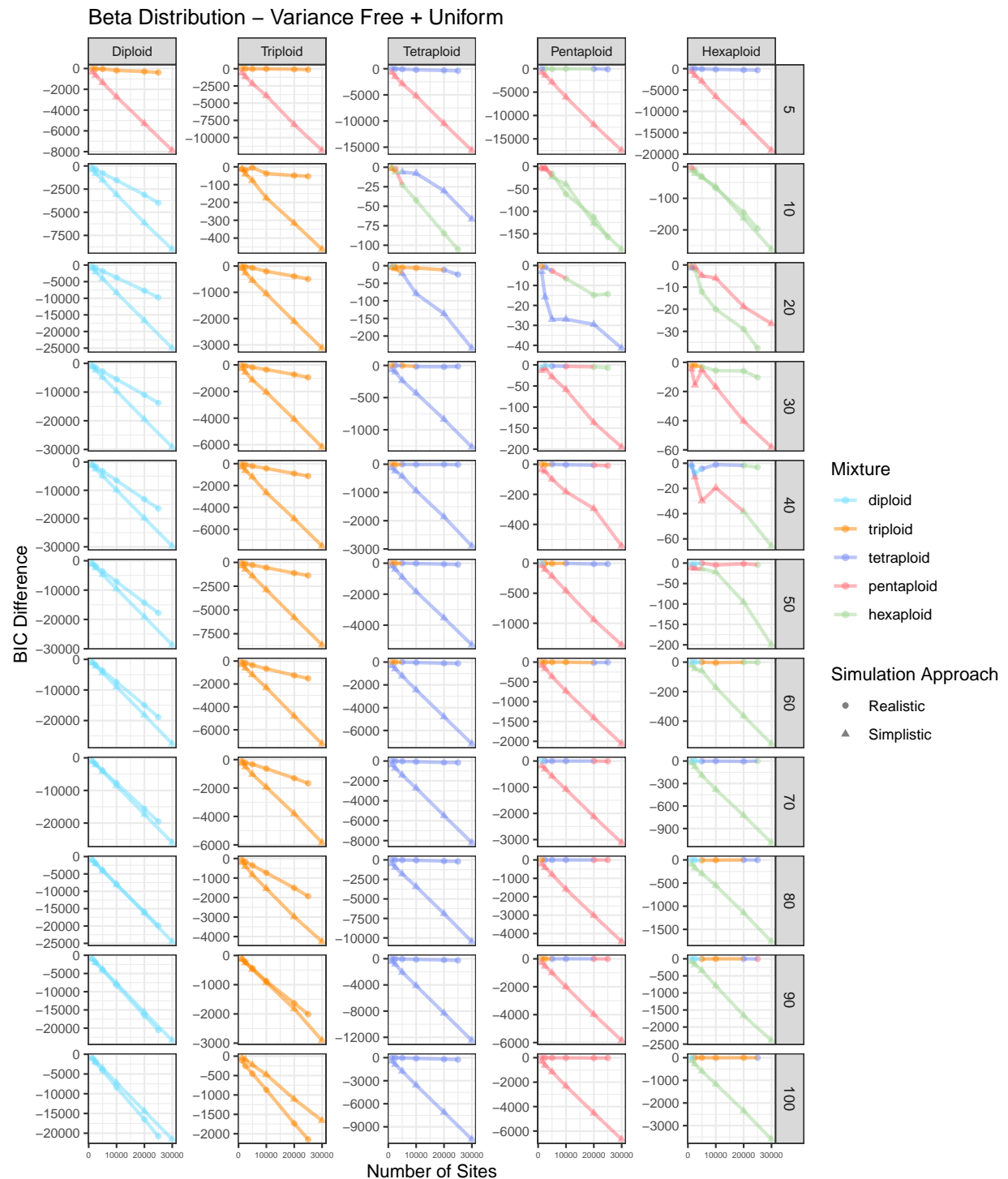


Figure S13. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with variance free with a uniform mixture.

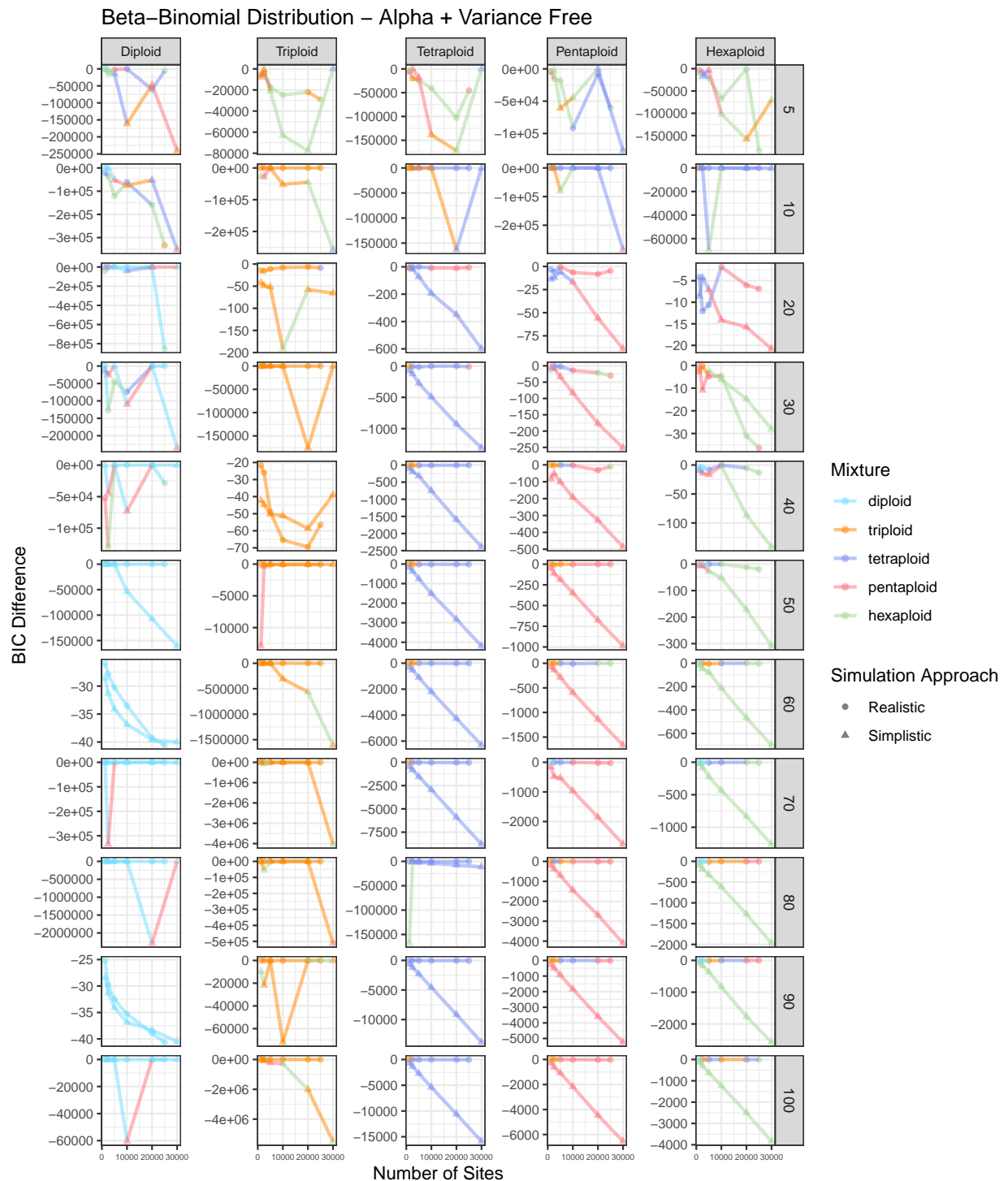


Figure S14. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha and variance free.



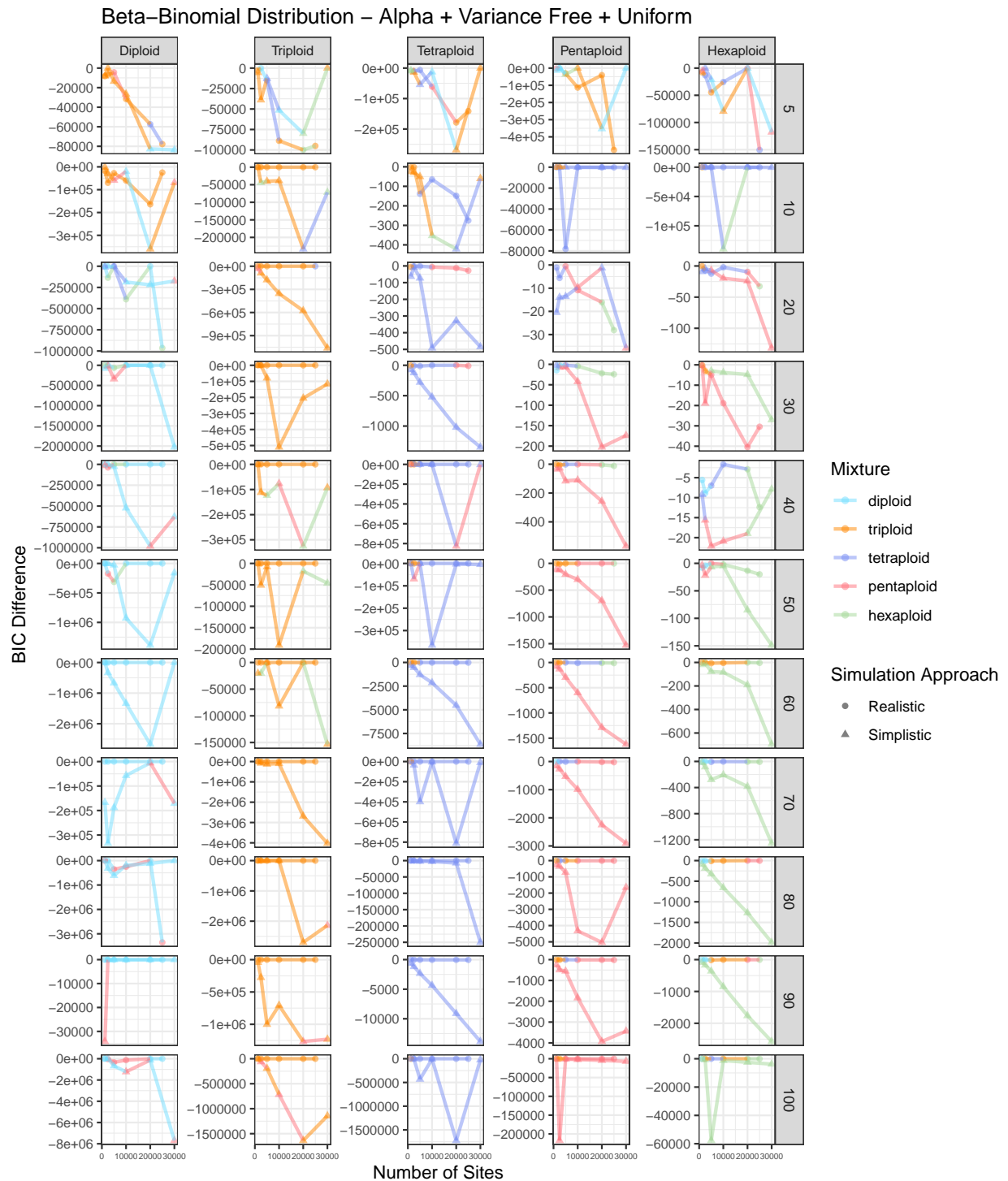


Figure S15. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha and variance free with a uniform mixture.



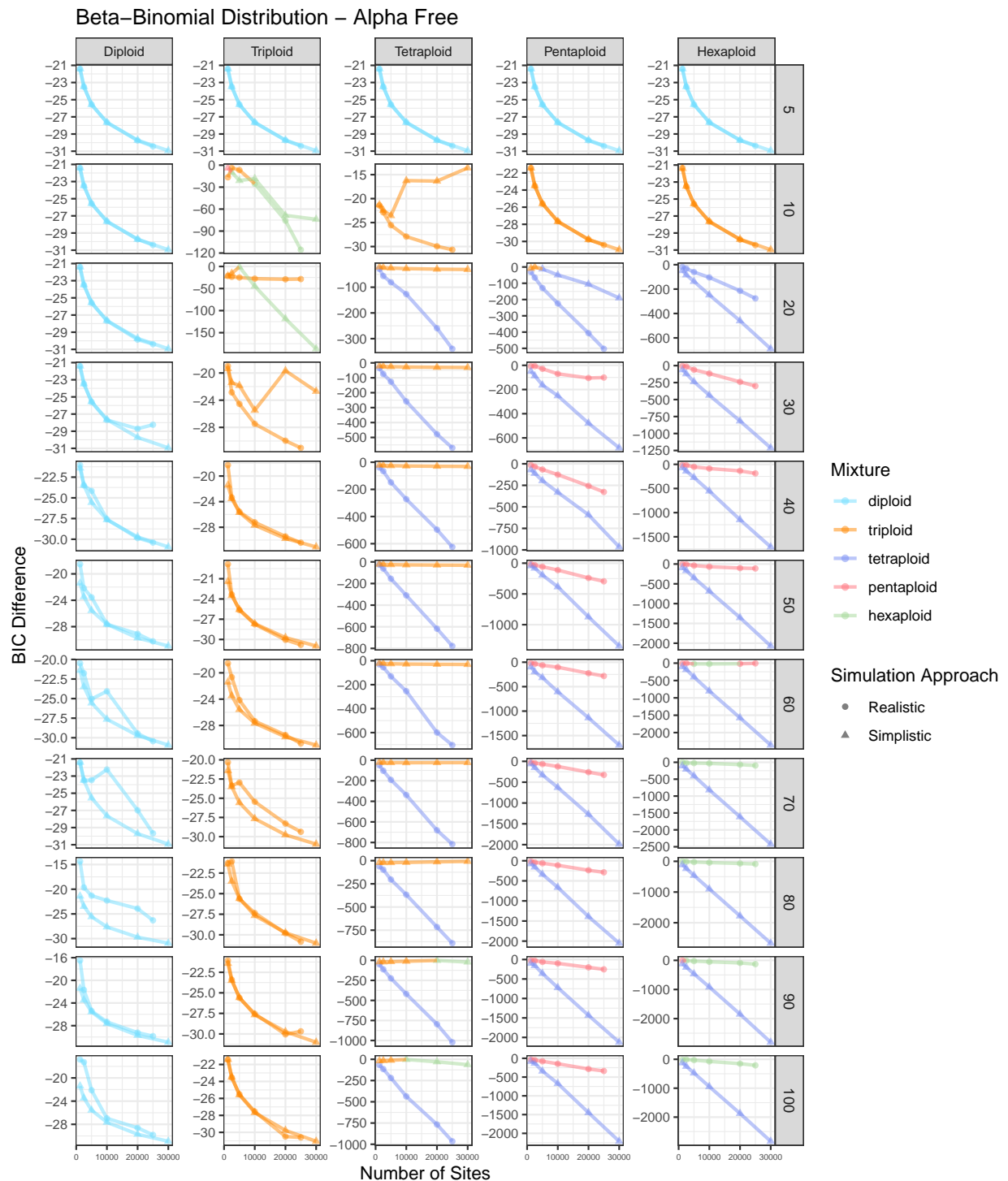


Figure S16. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha free.

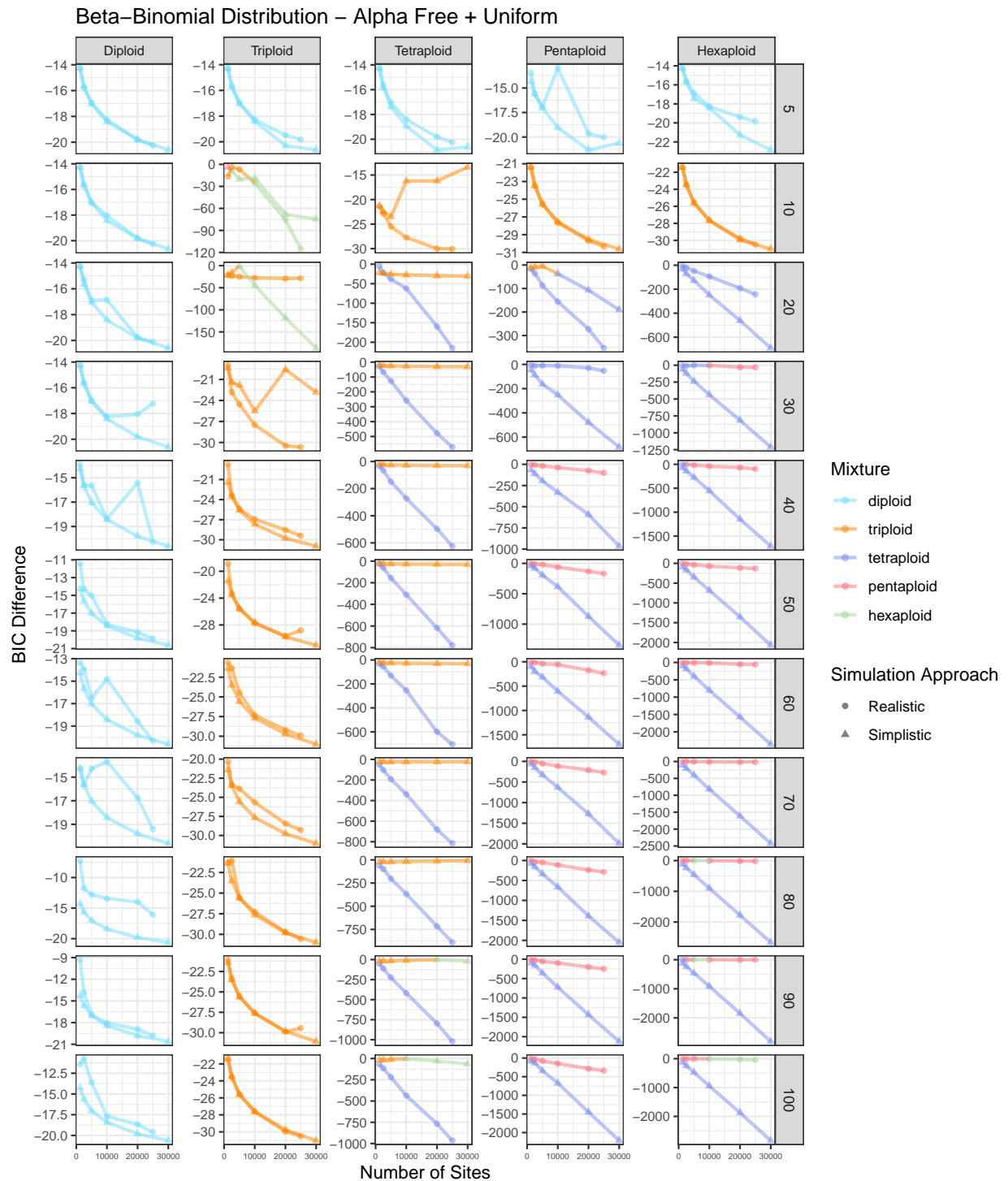


Figure S17. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha free with a uniform mixture.

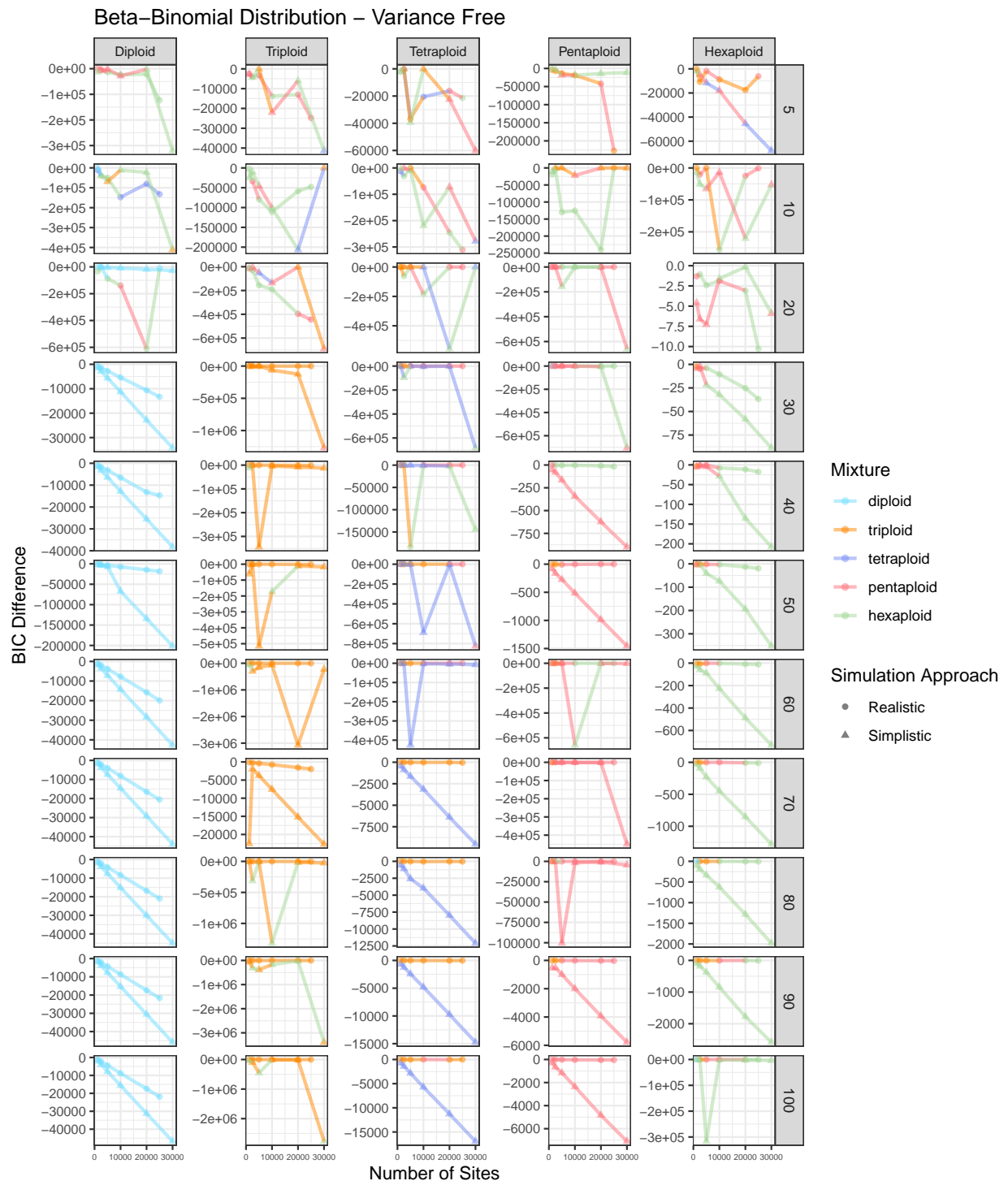


Figure S18. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with variance free.

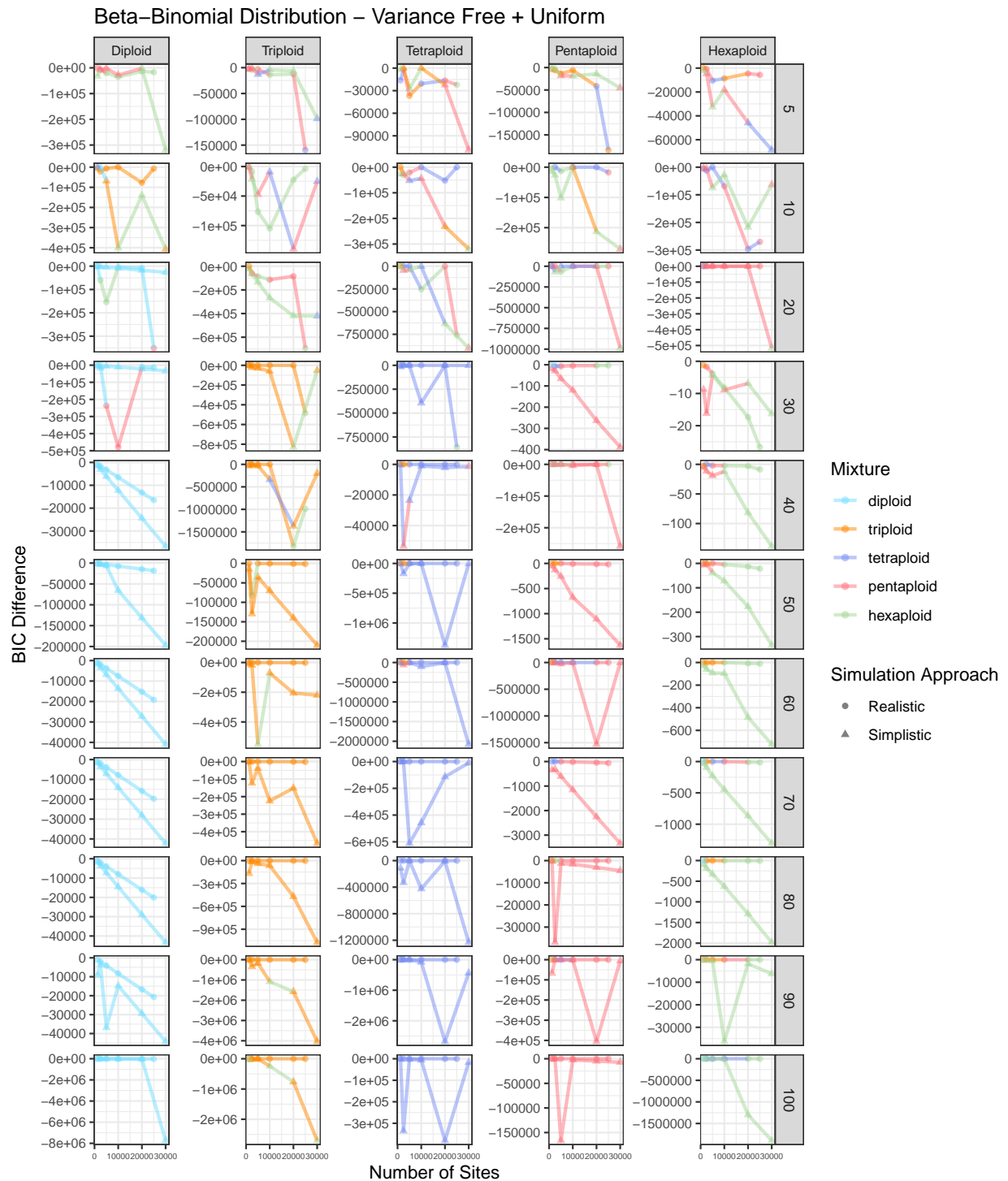


Figure S19. BIC score difference between the best and second best model for simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with variance free with a uniform mixture.

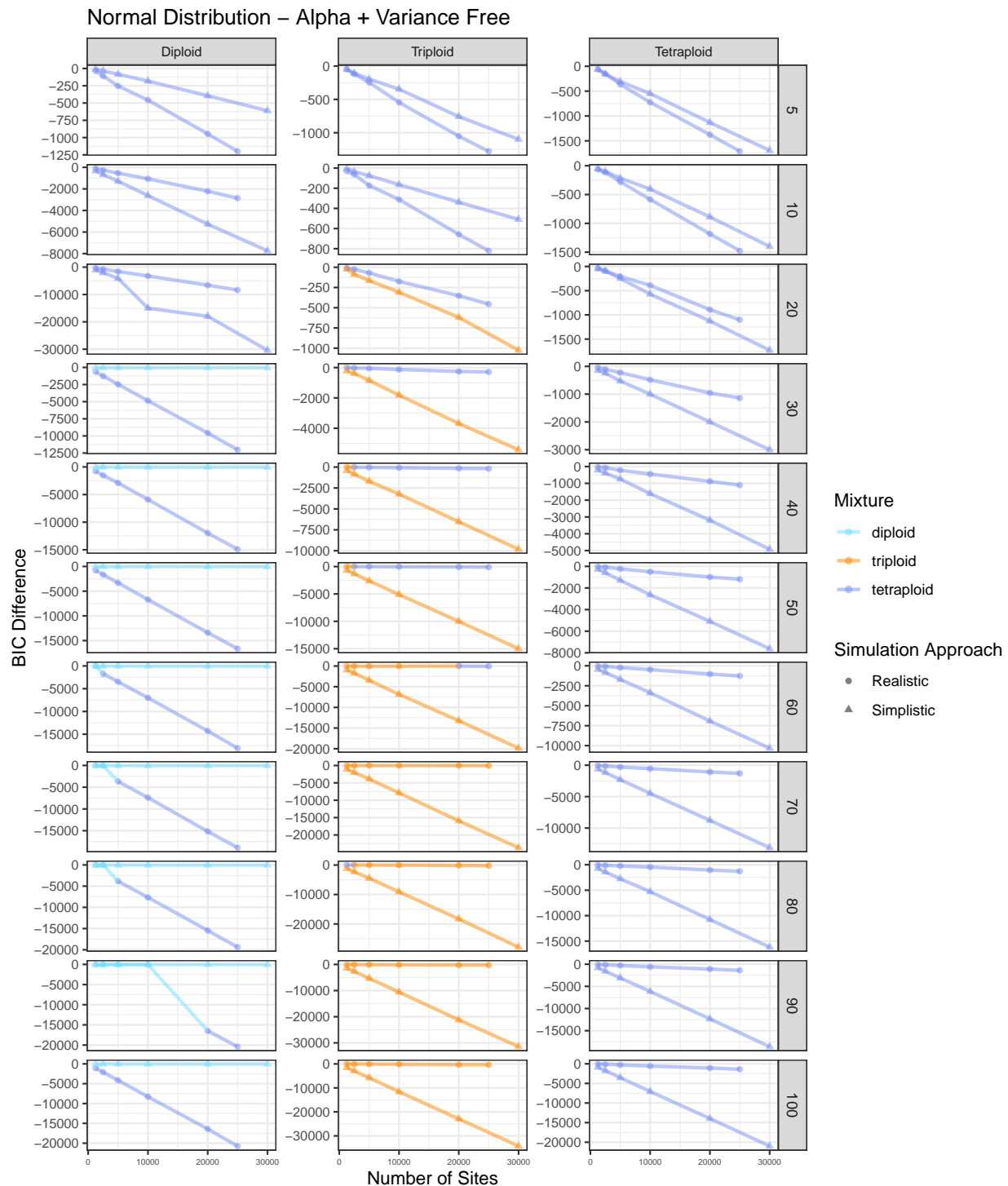


Figure S20. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha and variance free.

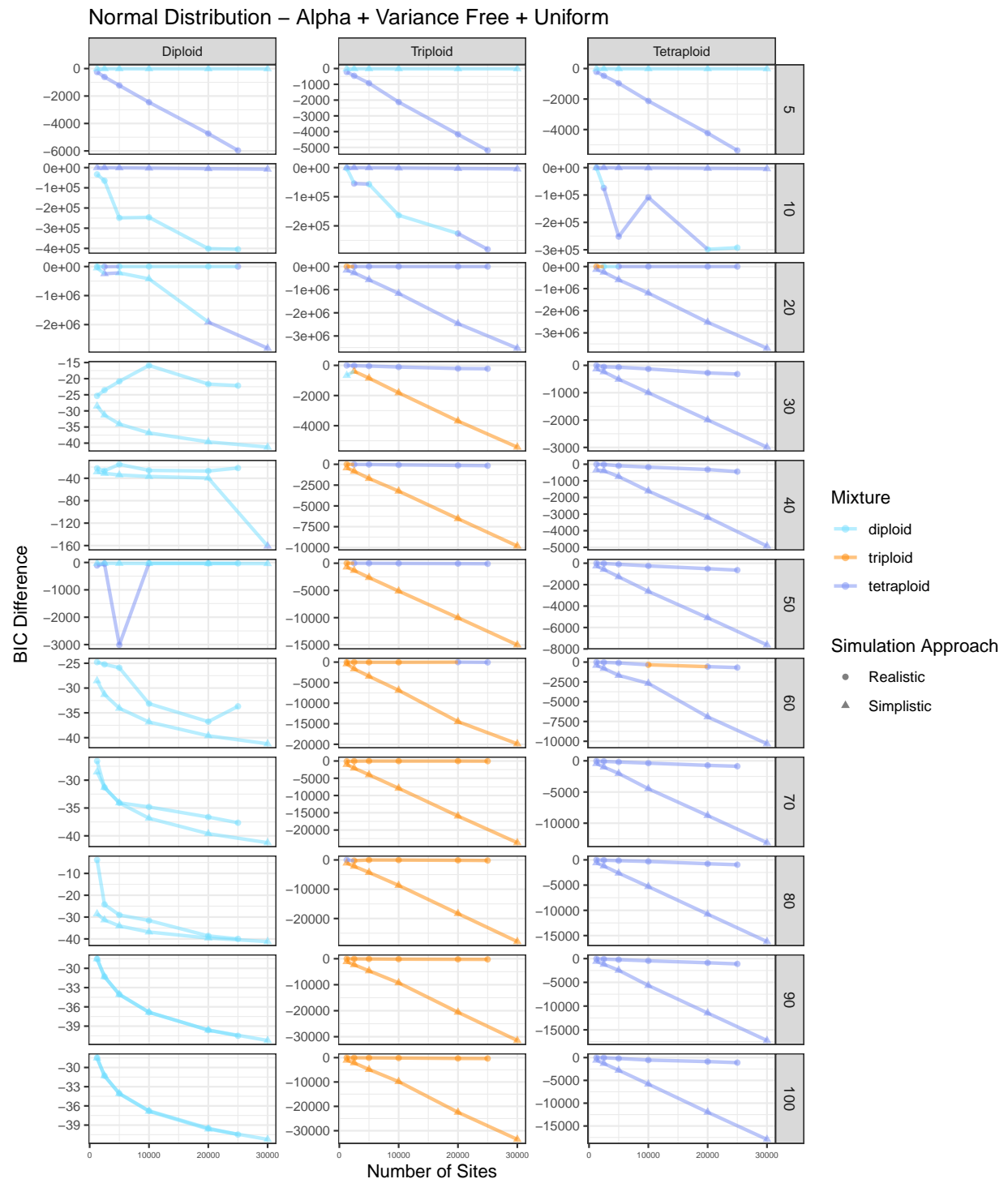


Figure S21. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha and variance free with a uniform mixture.

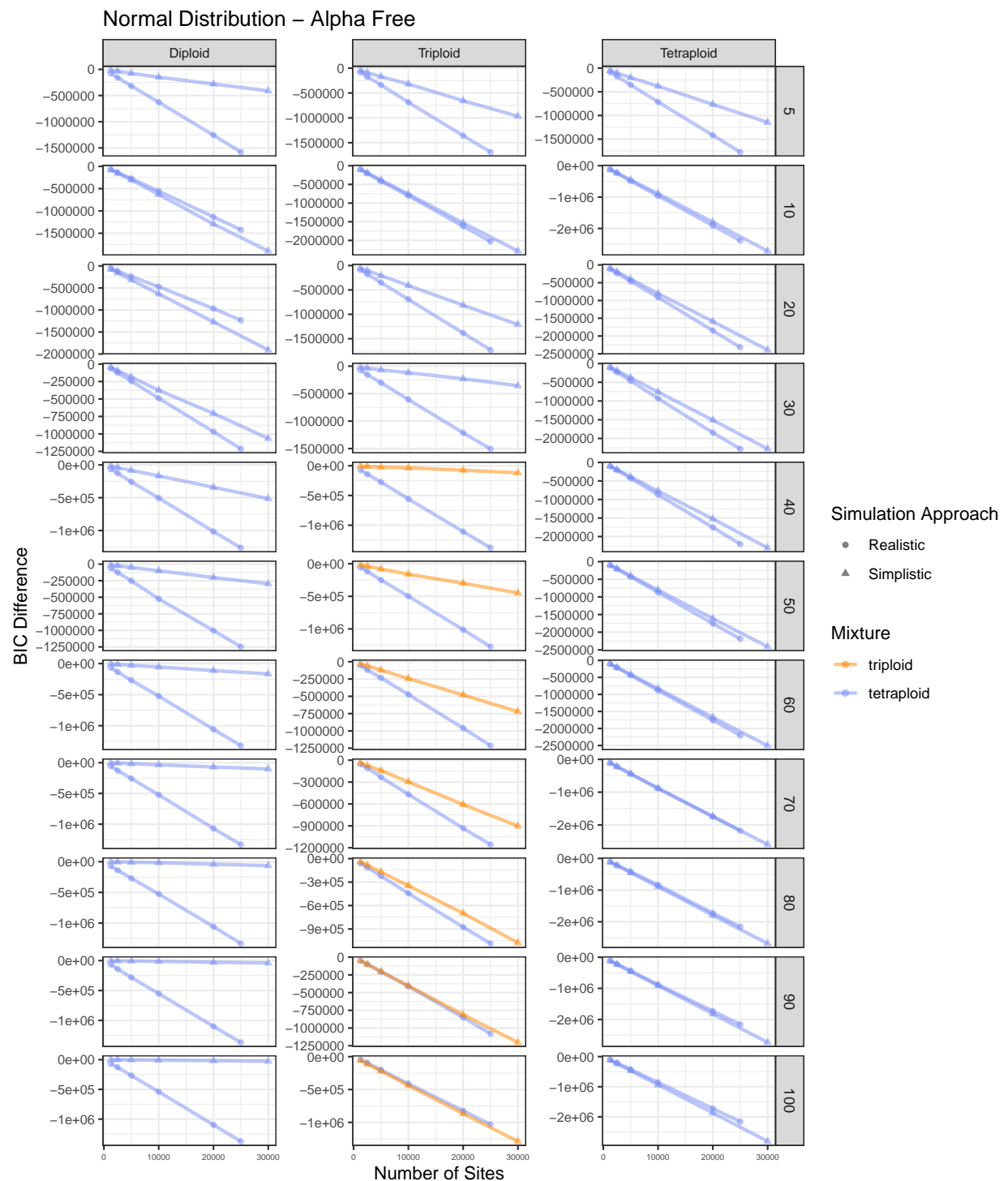


Figure S22. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha free.



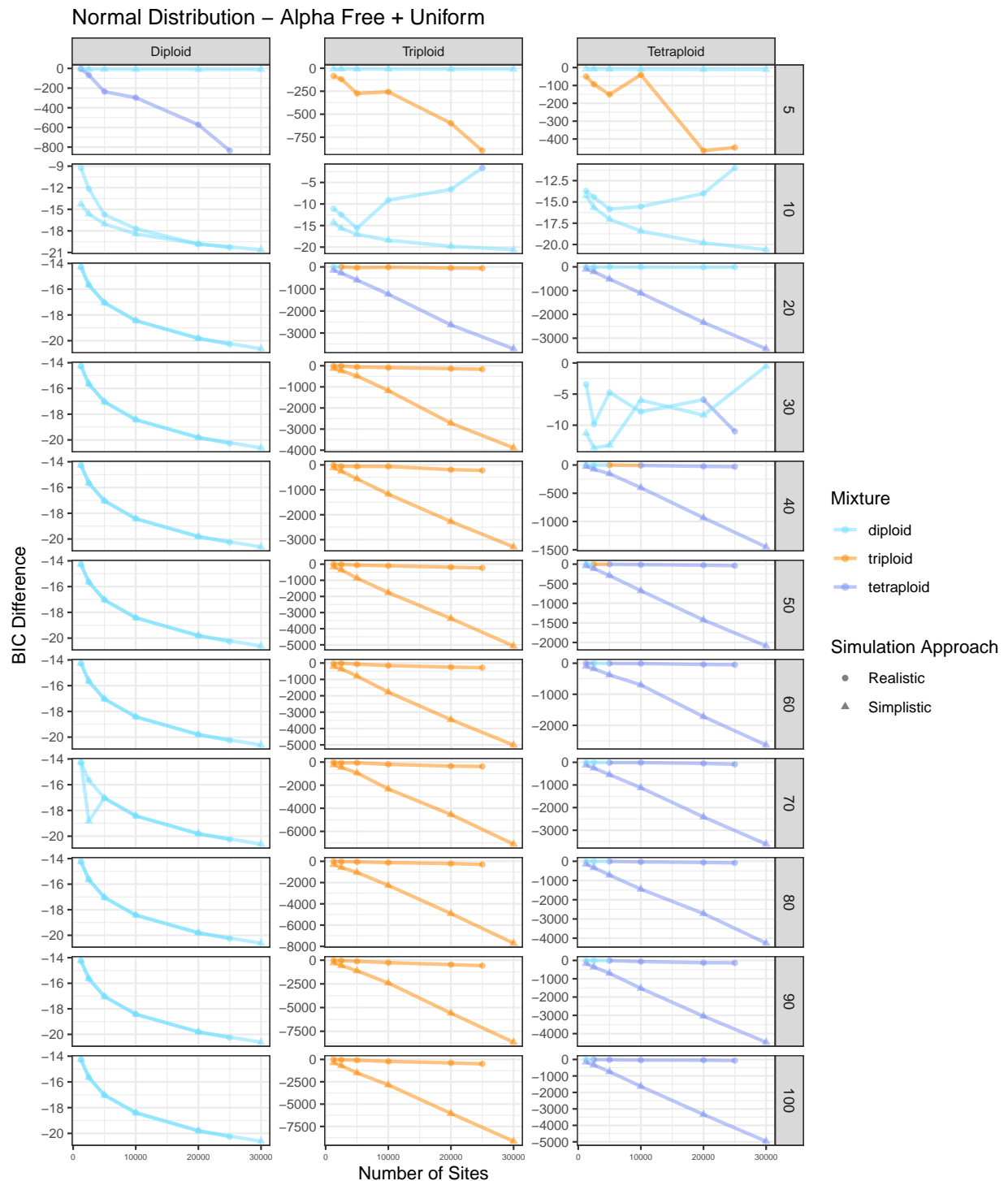


Figure S23. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha free with a uniform mixture.



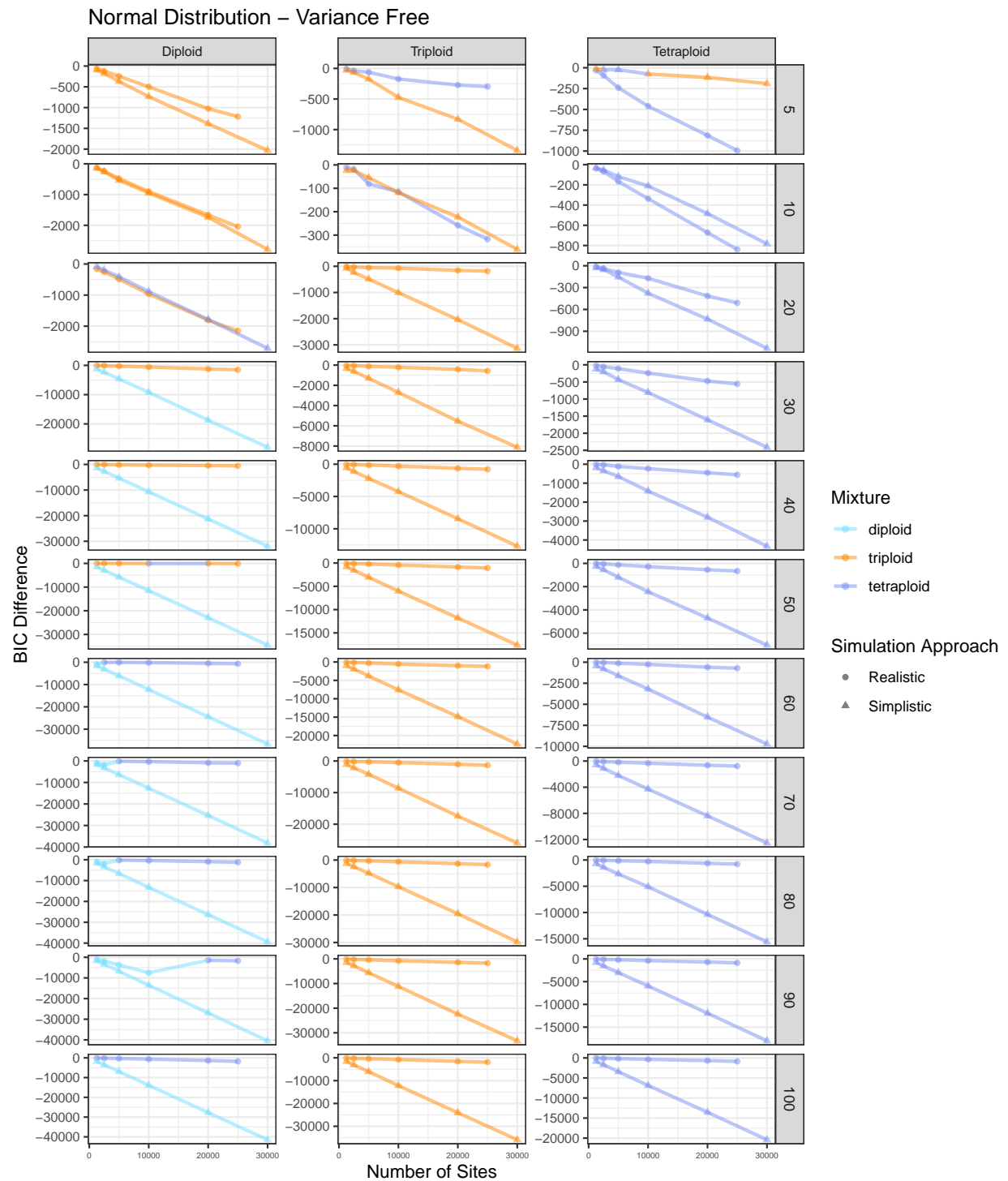


Figure S24. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with variance free.

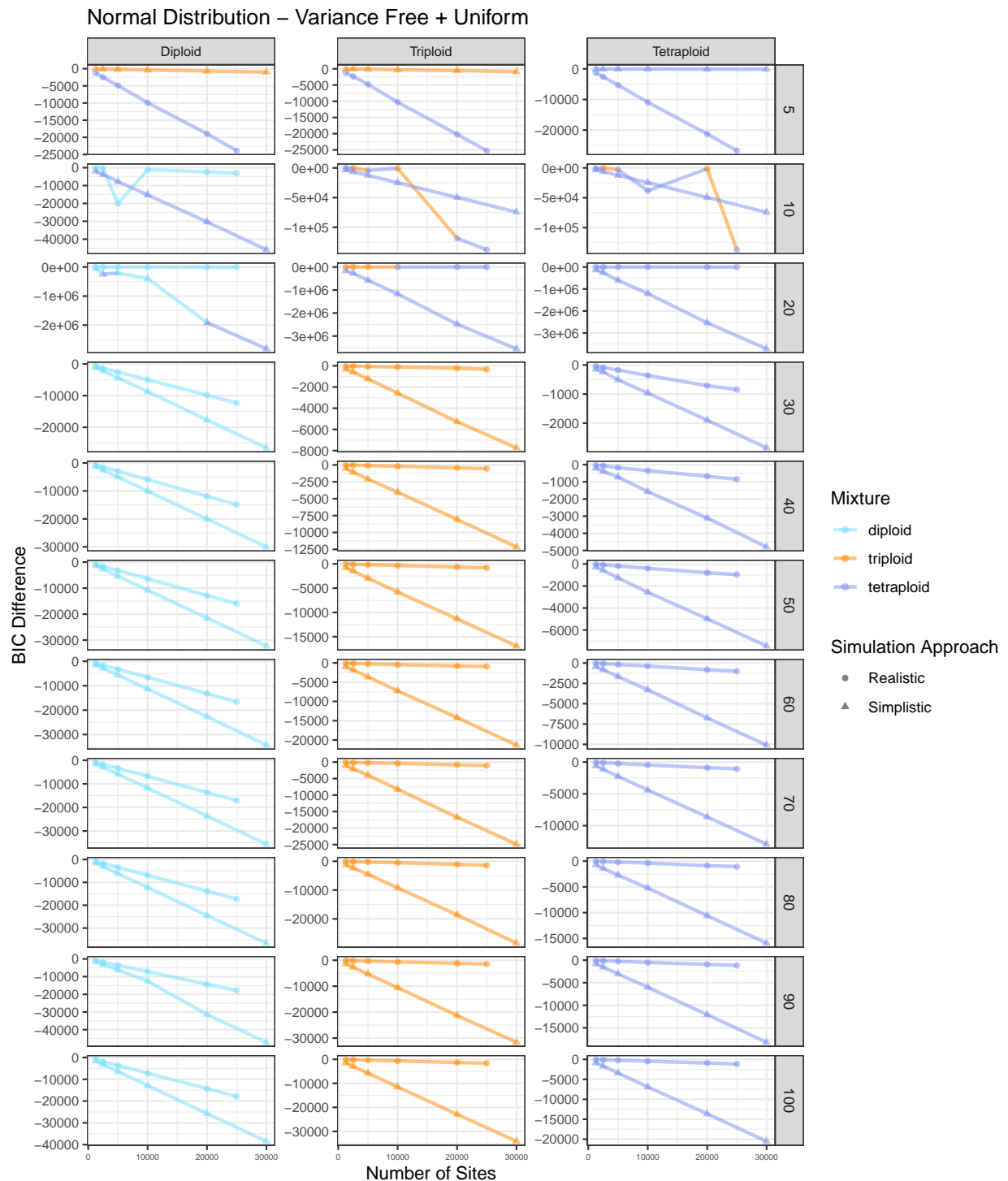


Figure S25. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with variance free with a uniform mixture.

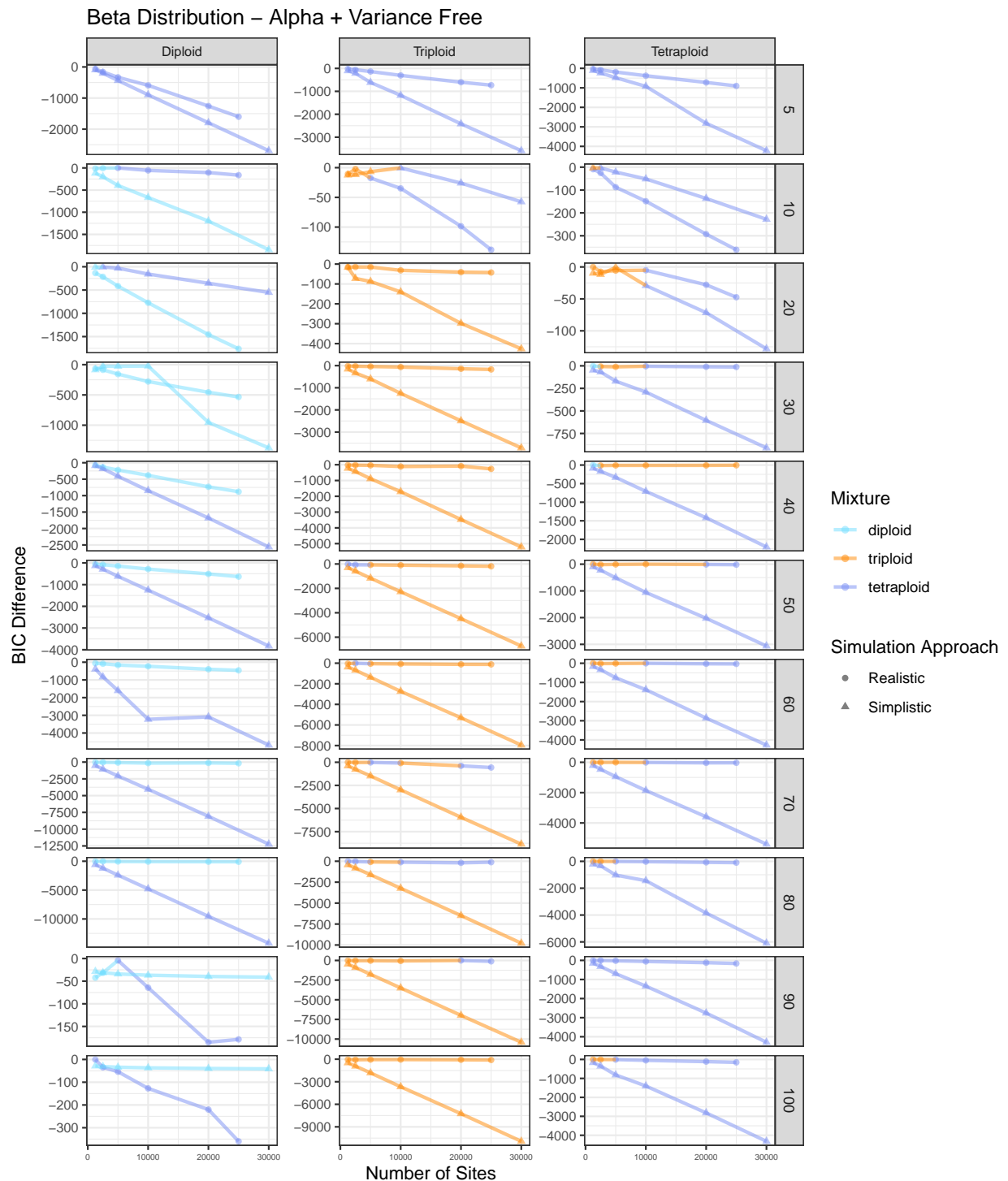


Figure S26. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha and variance free.

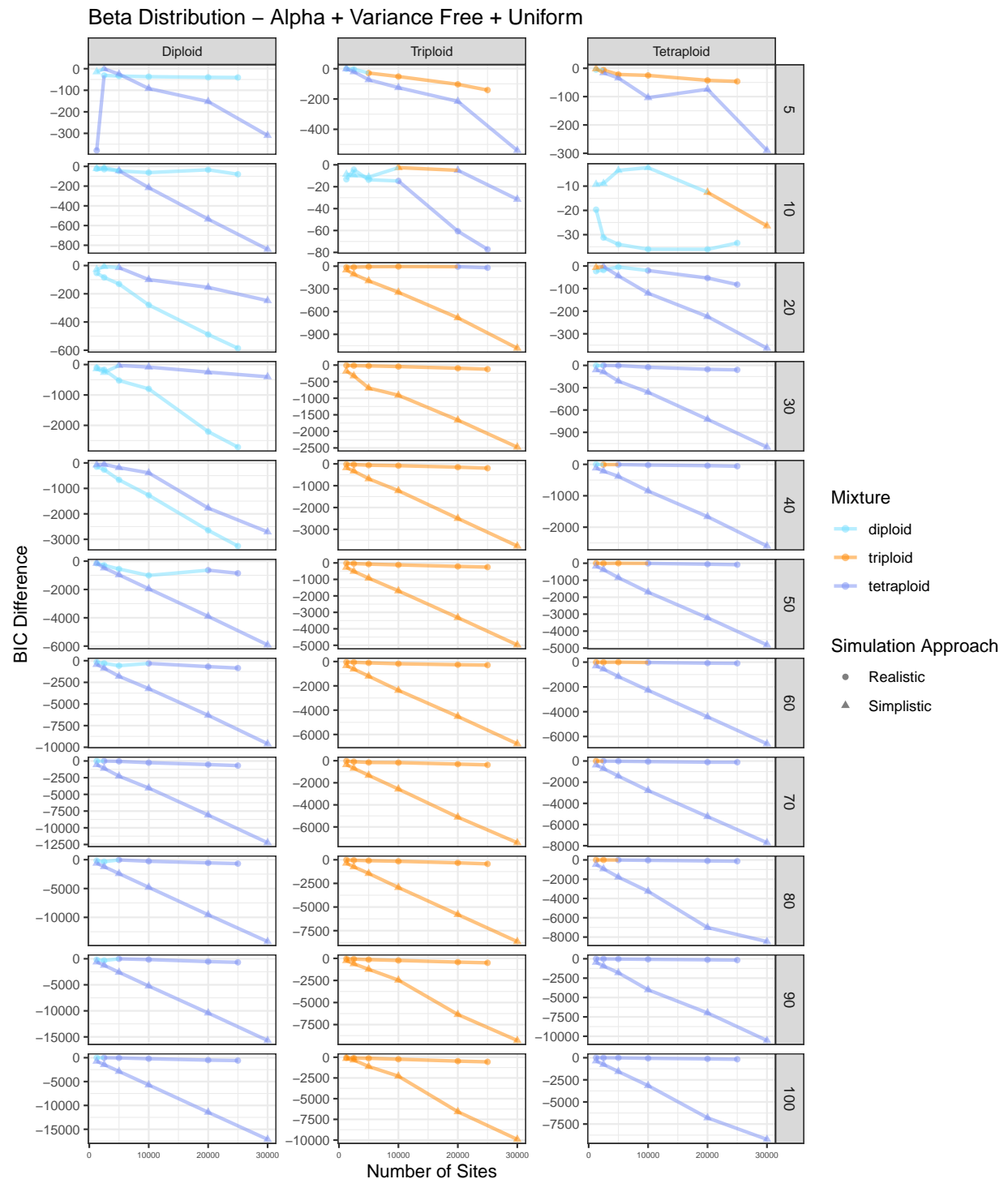


Figure S27. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha and variance free with a uniform mixture.

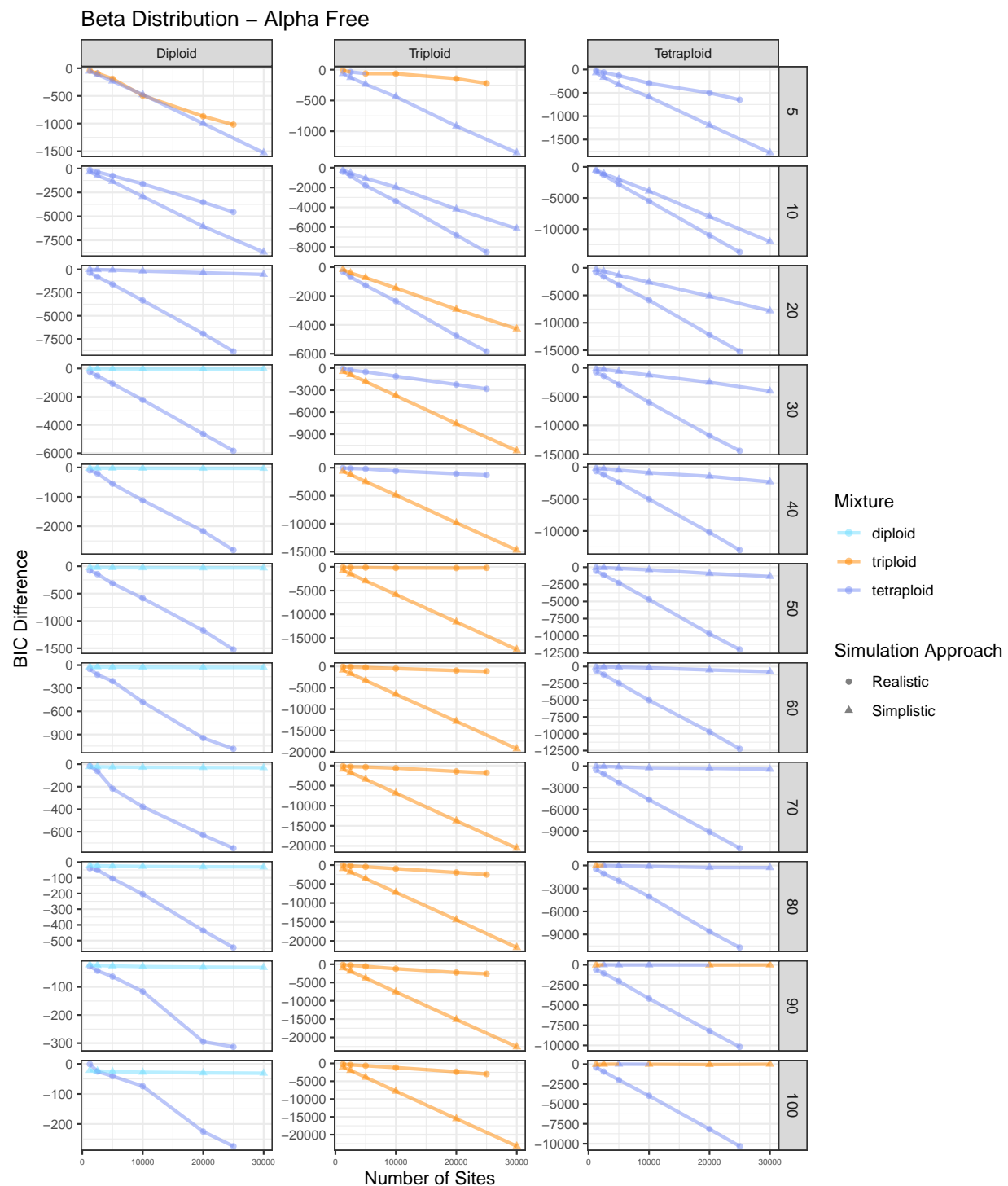


Figure S28. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha free.

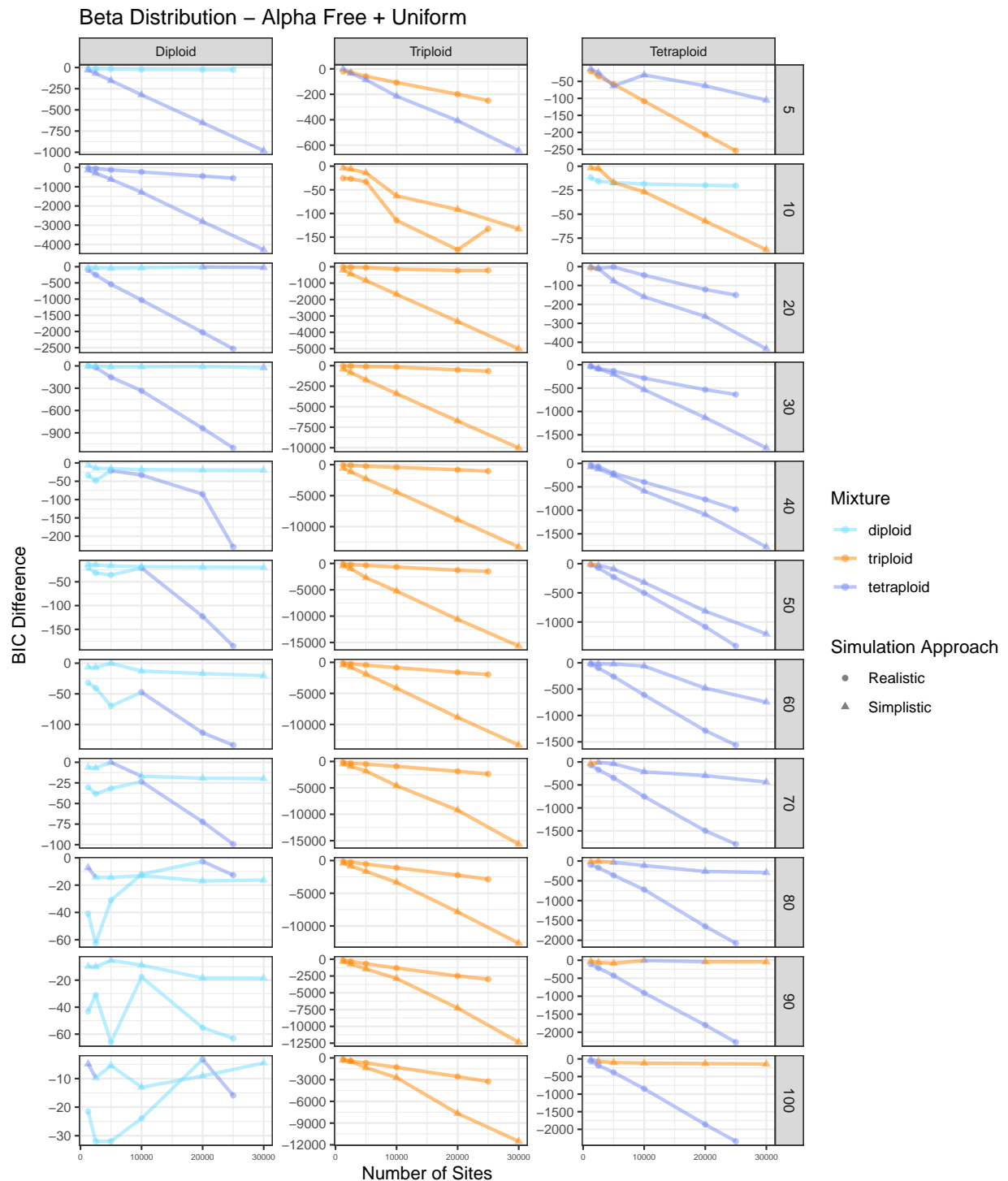


Figure S29. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha free with a uniform mixture.

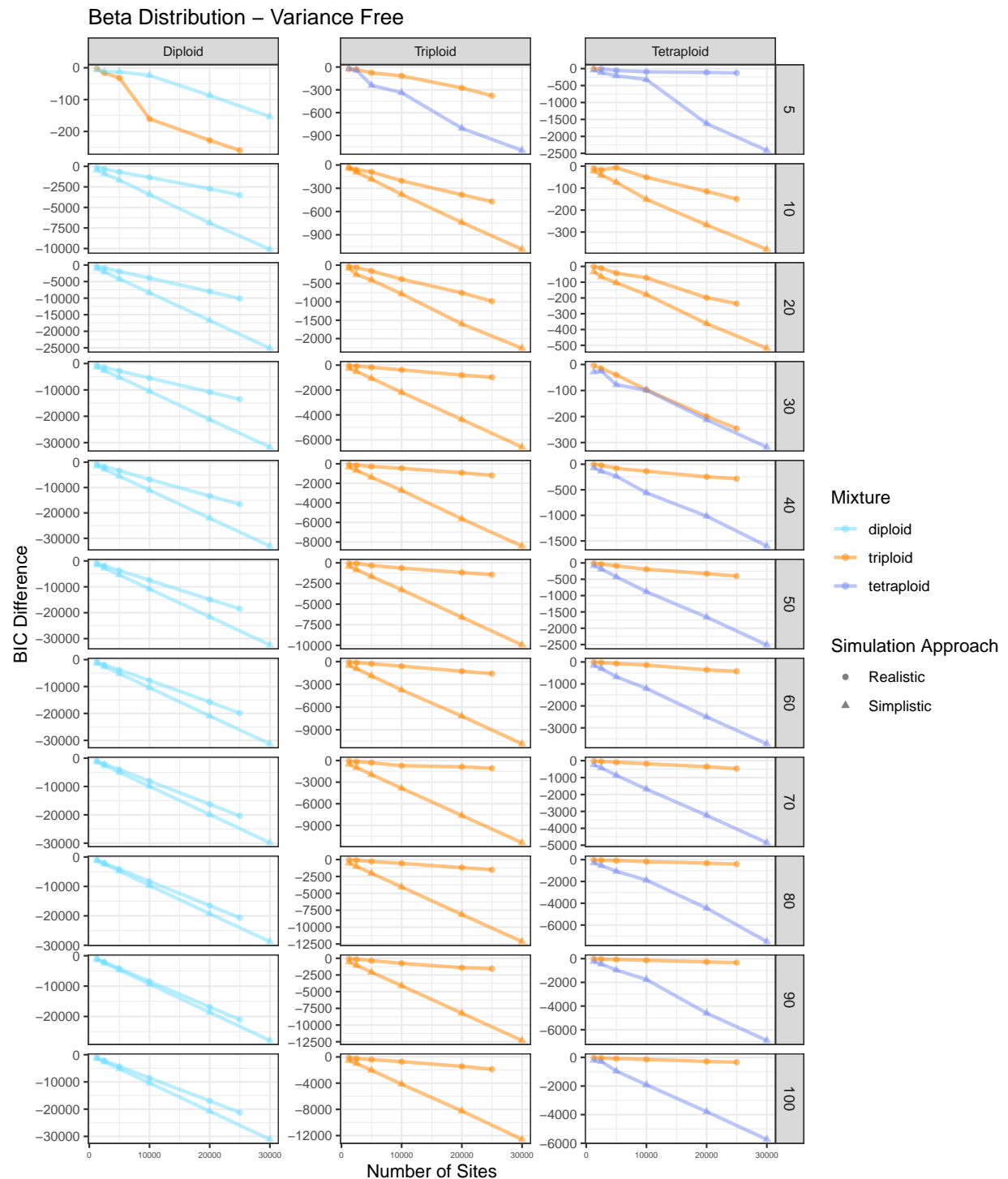


Figure S30. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with variance free.

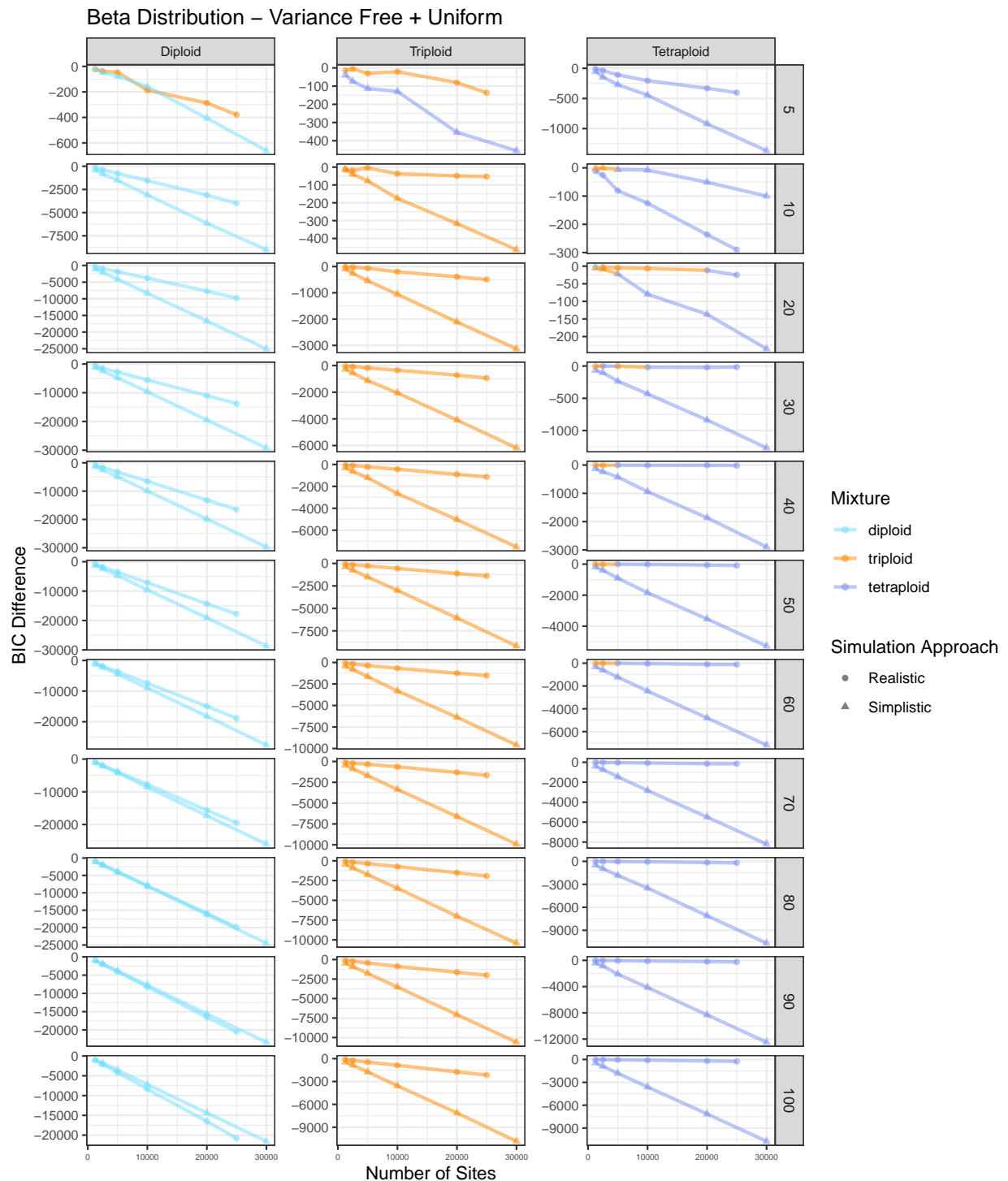


Figure S31. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with variance free with a uniform mixture.



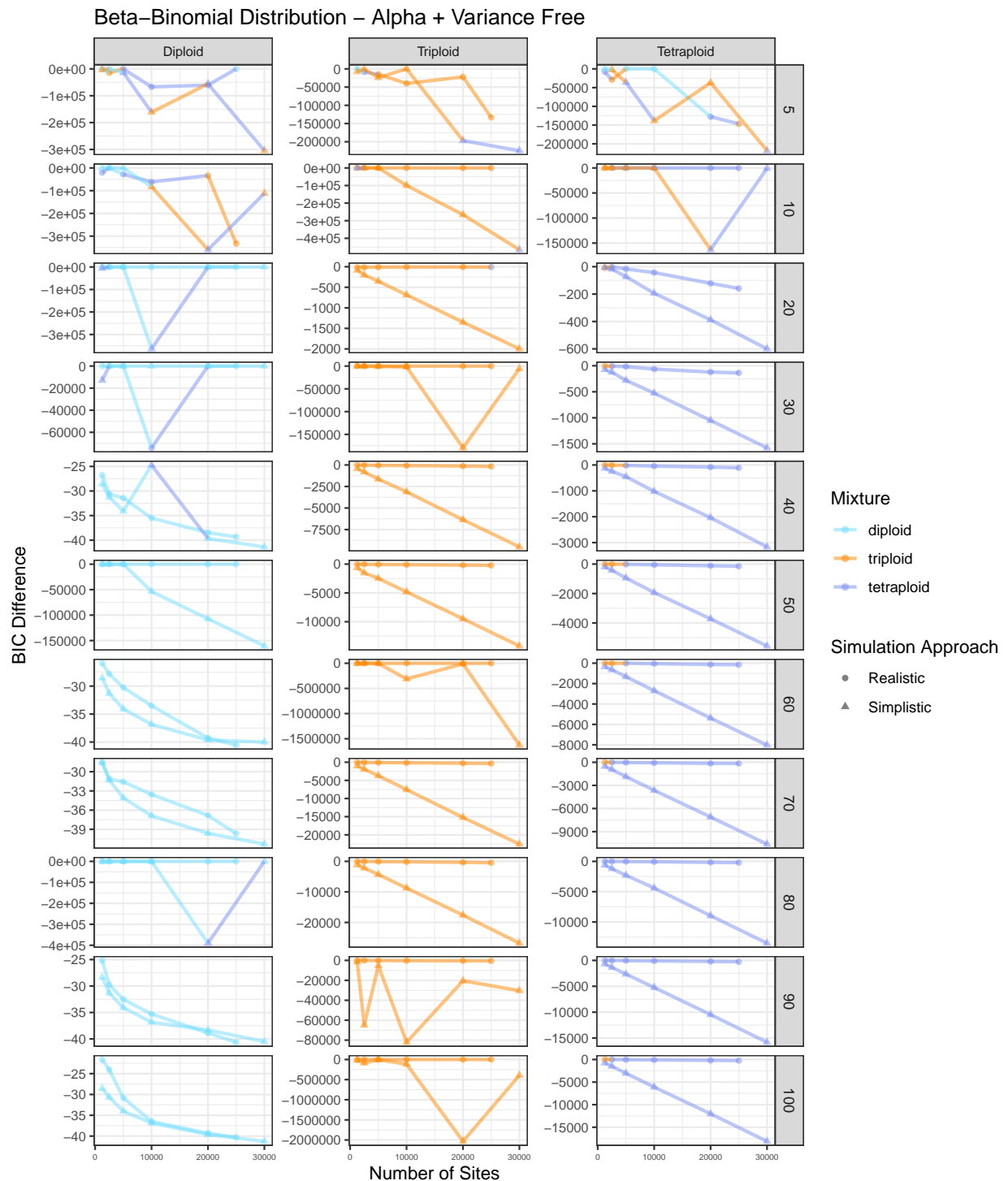


Figure S32. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha and variance free.

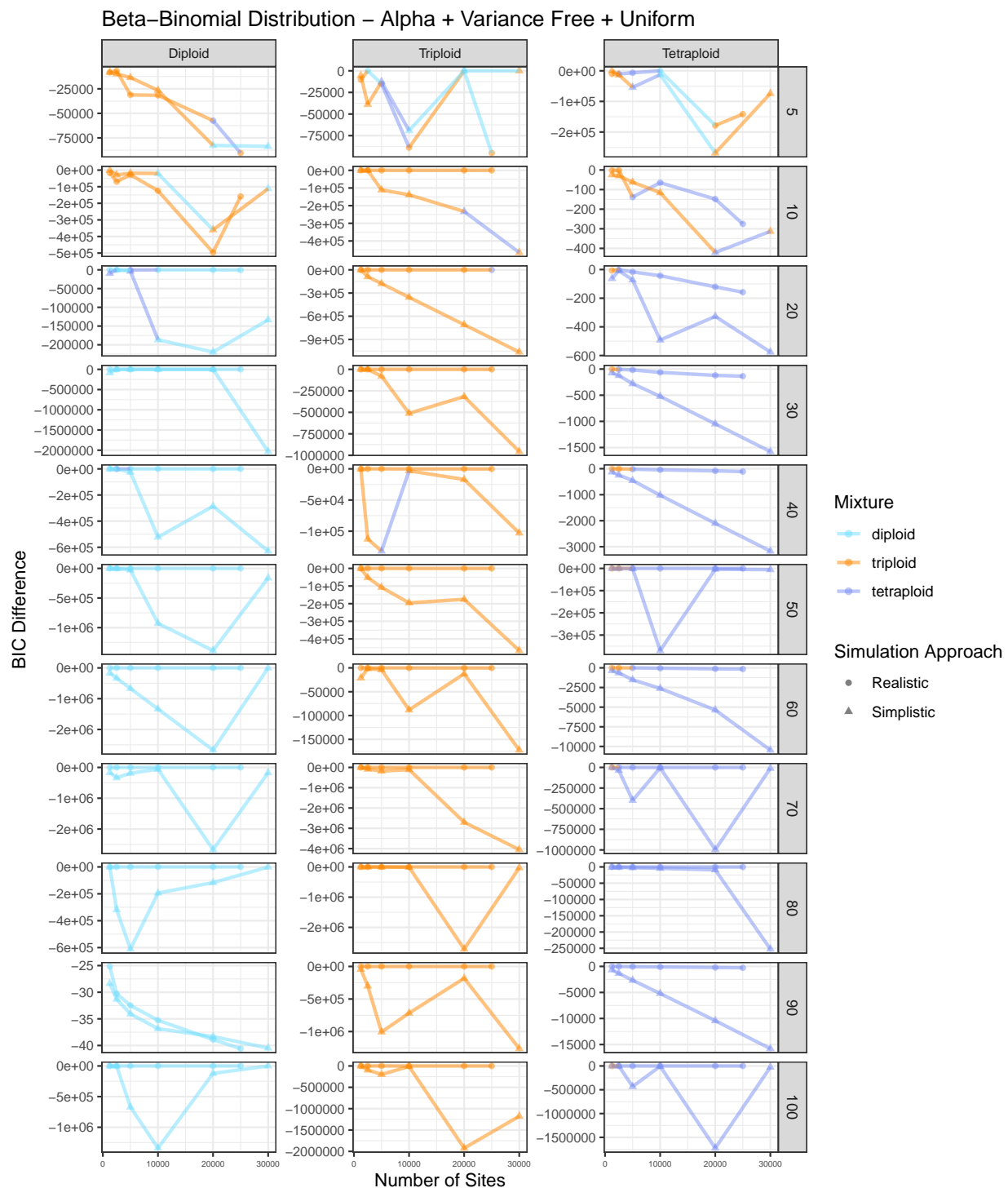


Figure S33. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha and variance free with a uniform mixture.

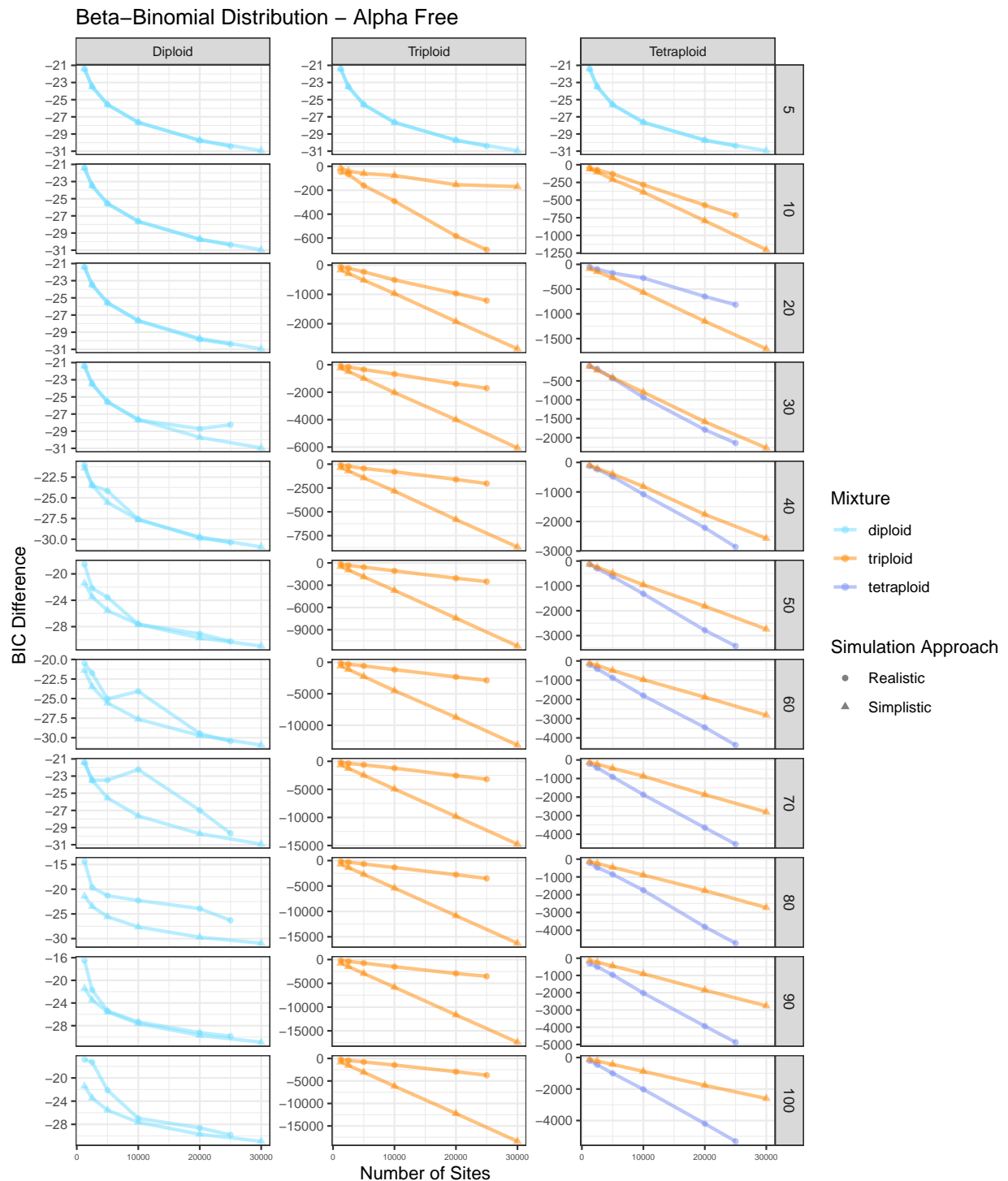


Figure S34. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha free.

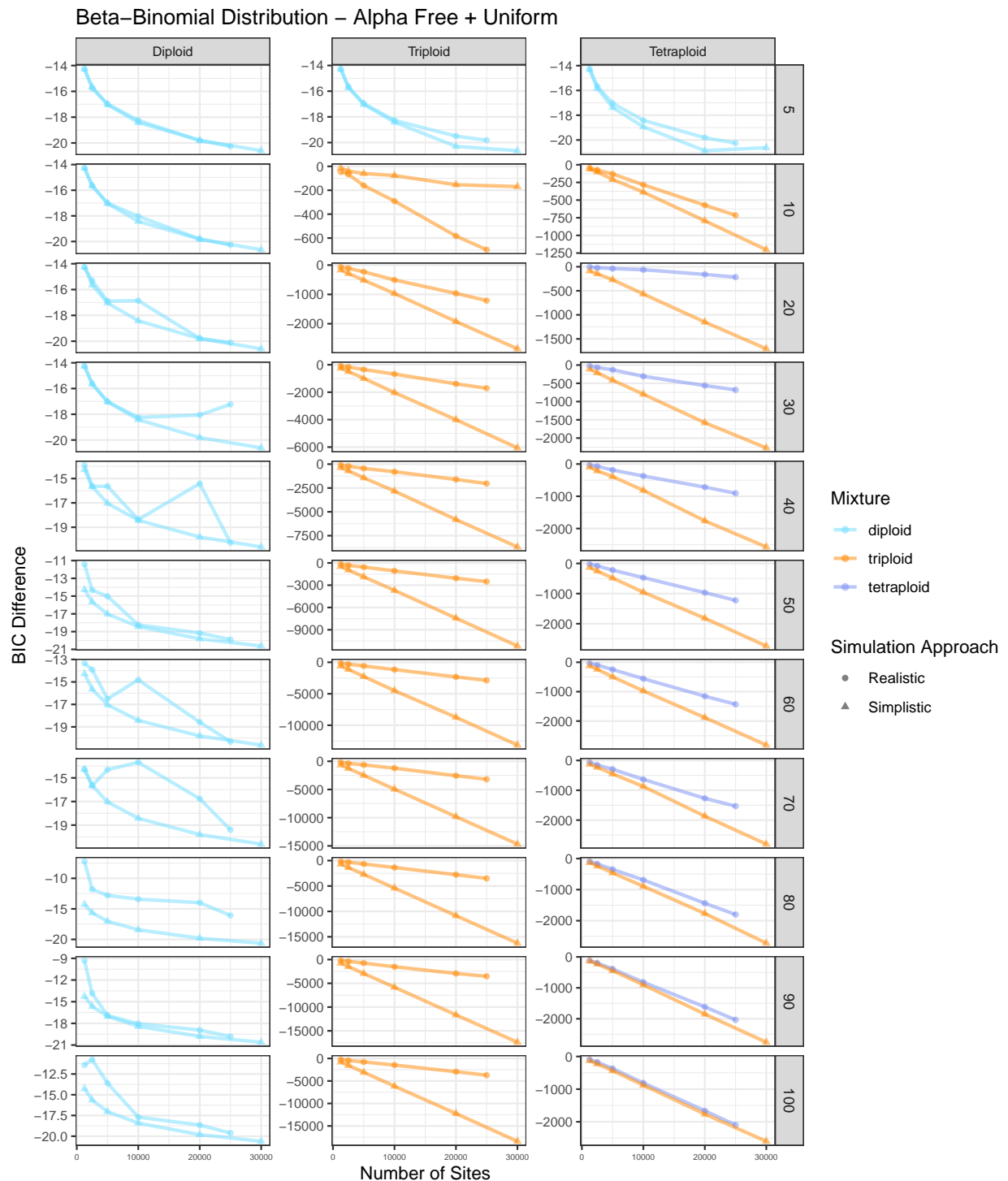


Figure S35. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha free with a uniform mixture.

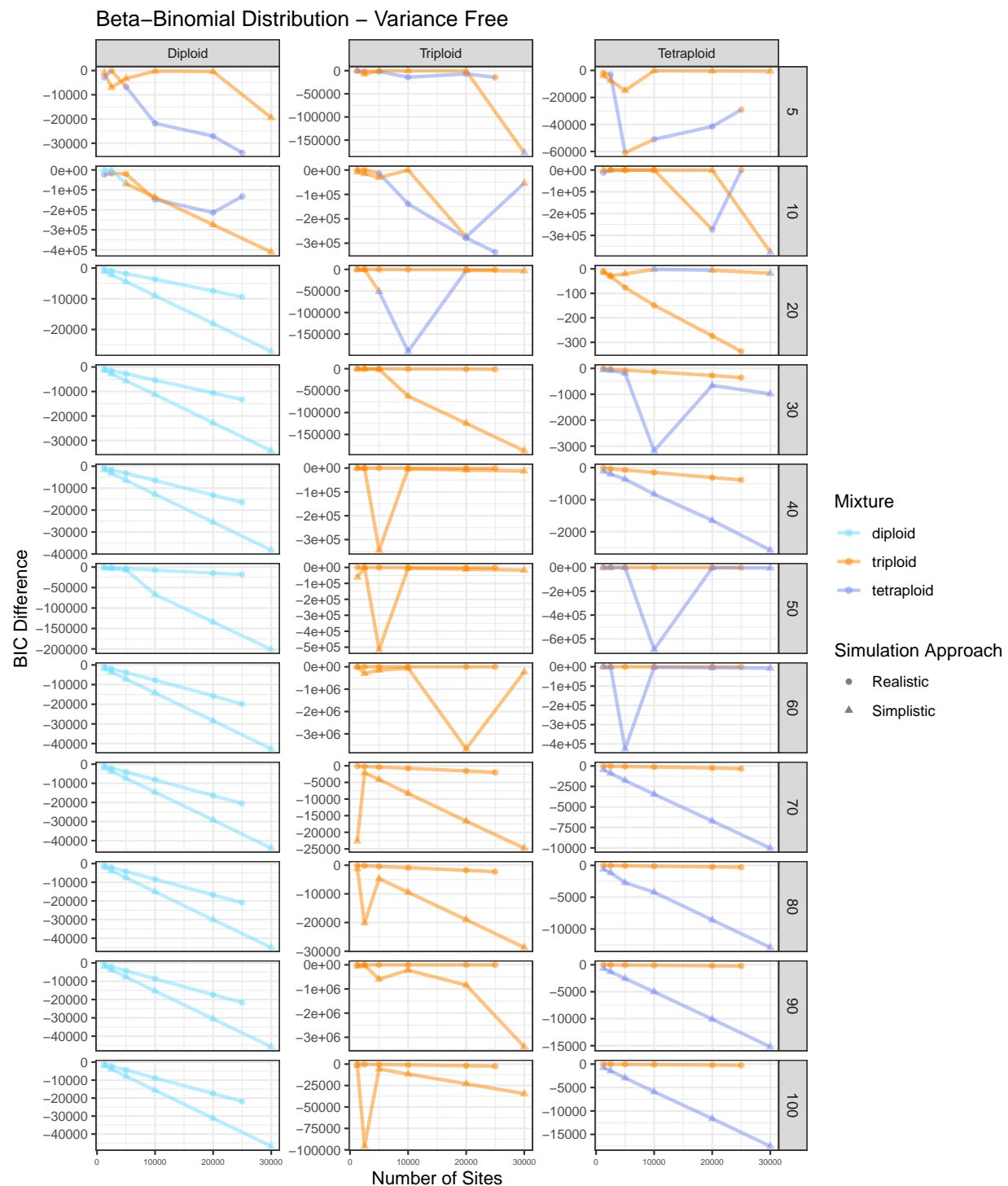


Figure S36. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with variance free.

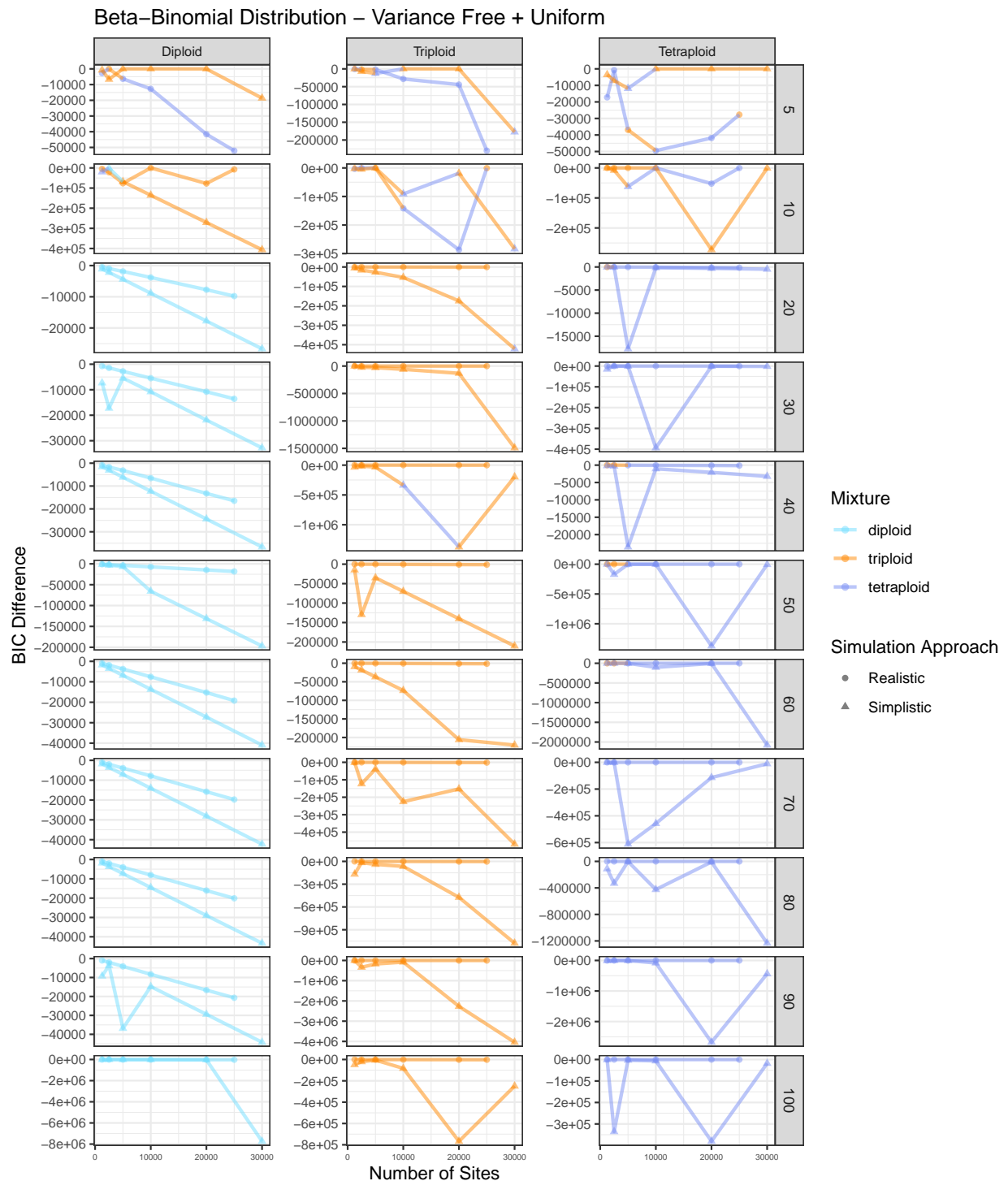


Figure S37. BIC score difference between the best and second best model for simulated diploid, triploid, and tetraploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with variance free with a uniform mixture.

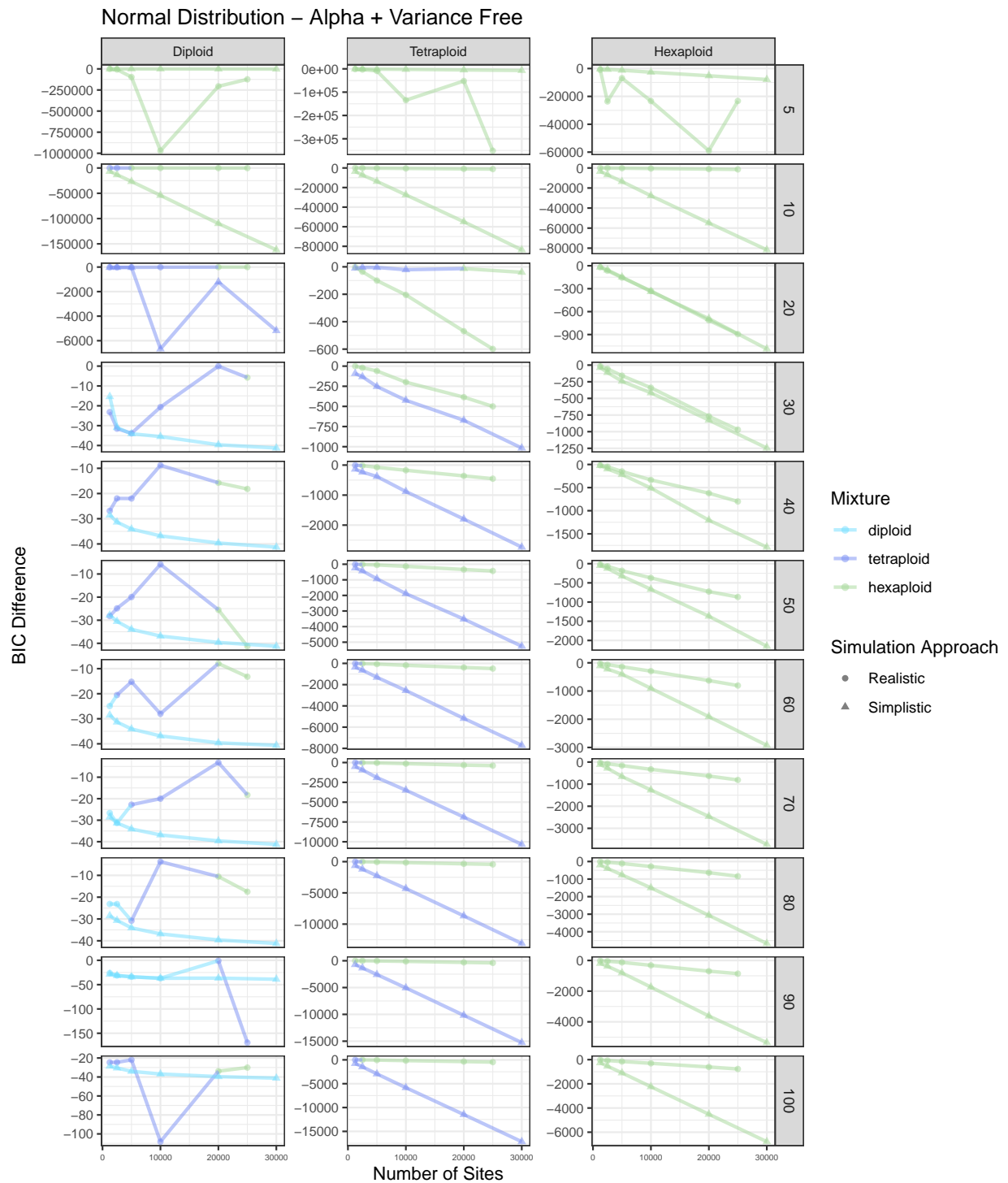


Figure S38. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha and variance free.

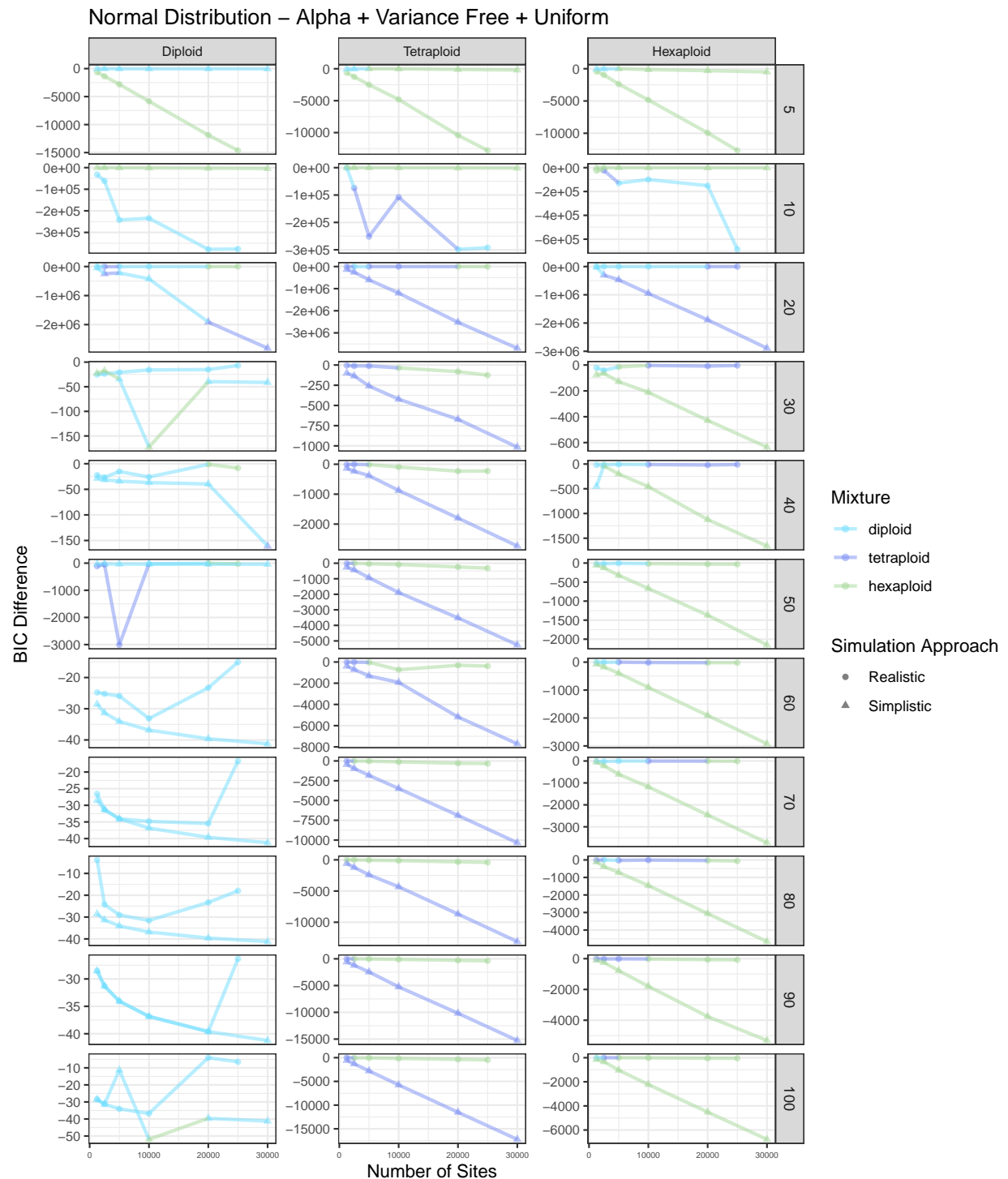


Figure S39. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha and variance free with a uniform mixture.



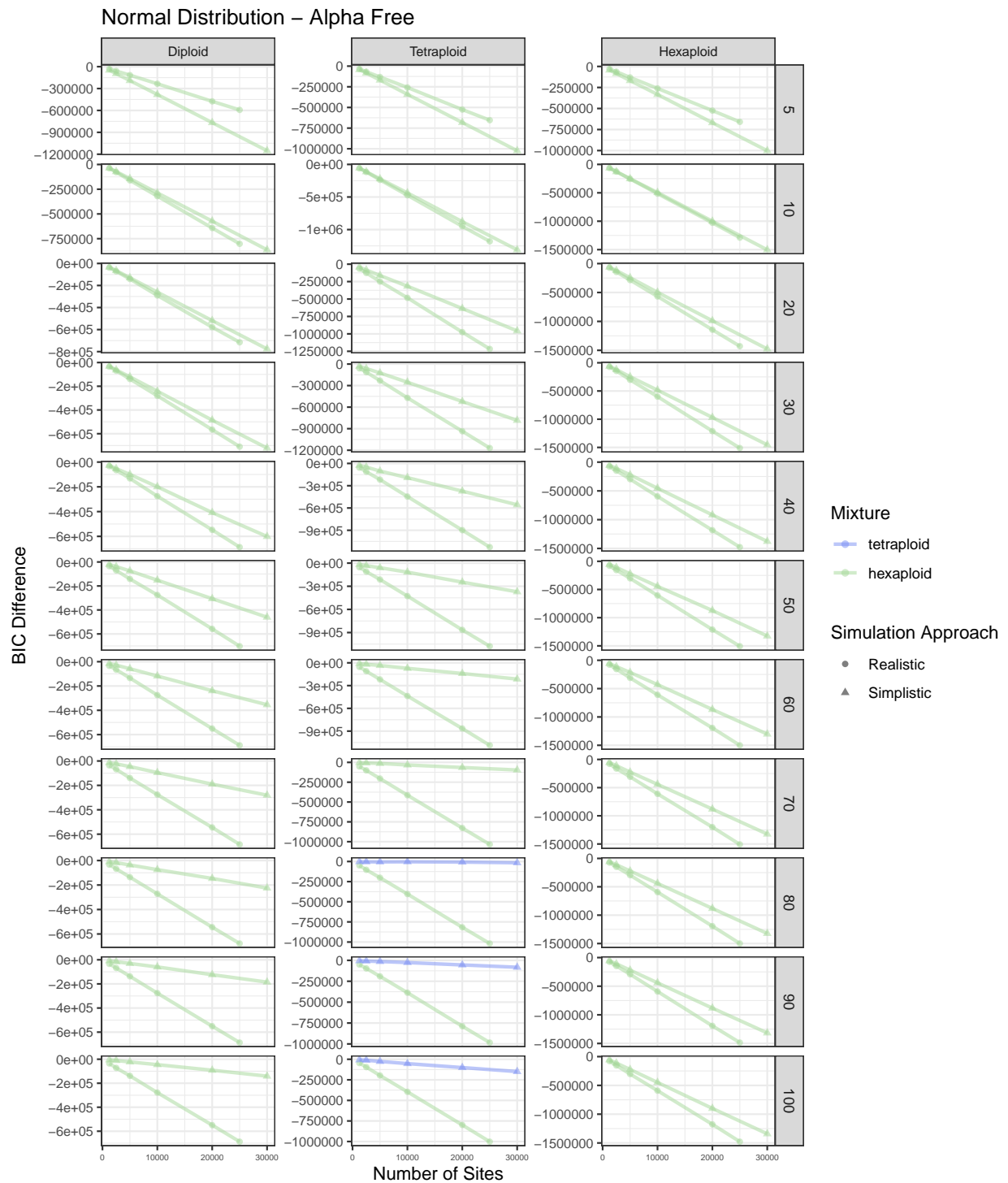


Figure S40. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha free.

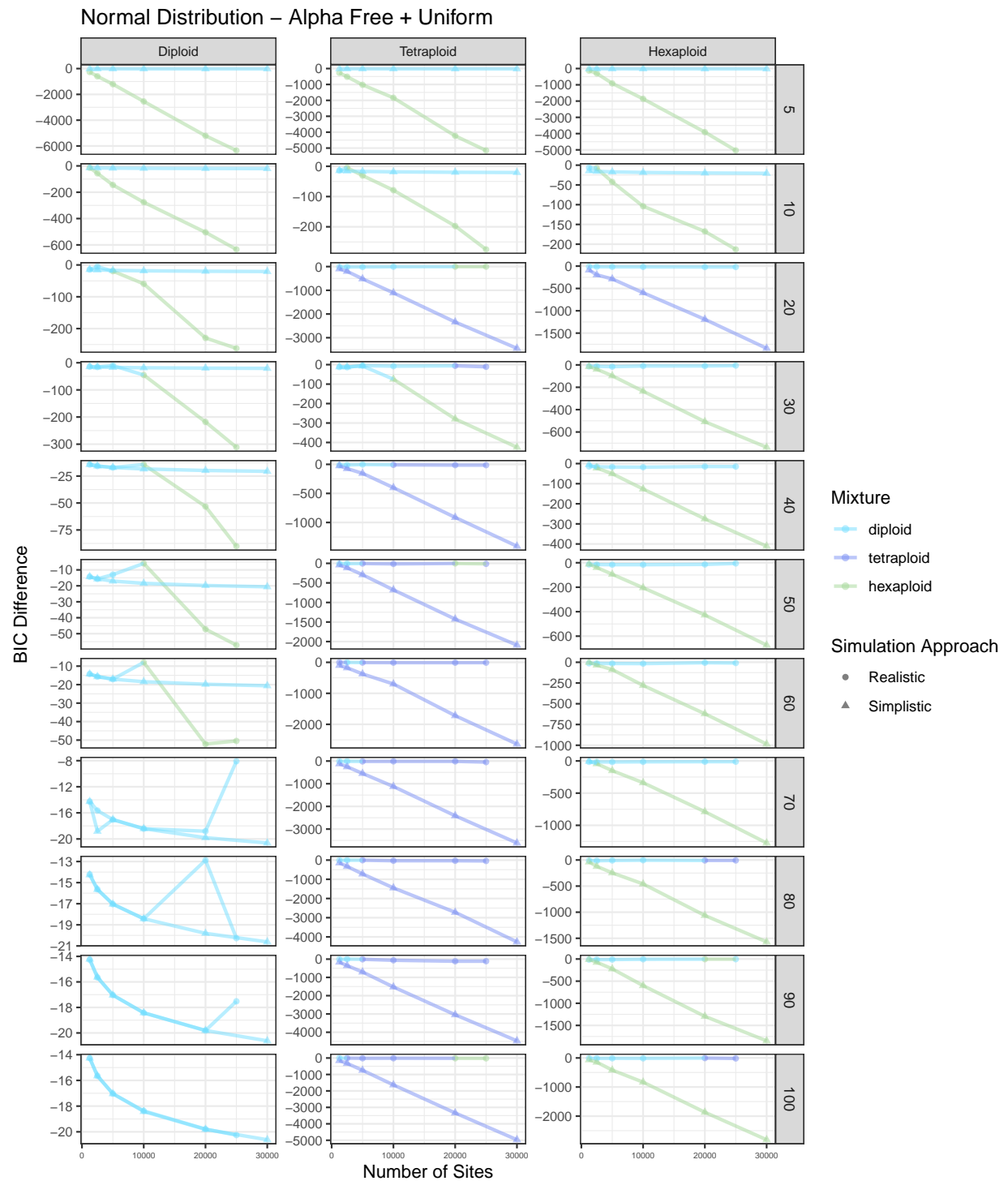


Figure S41. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with alpha free with a uniform mixture.

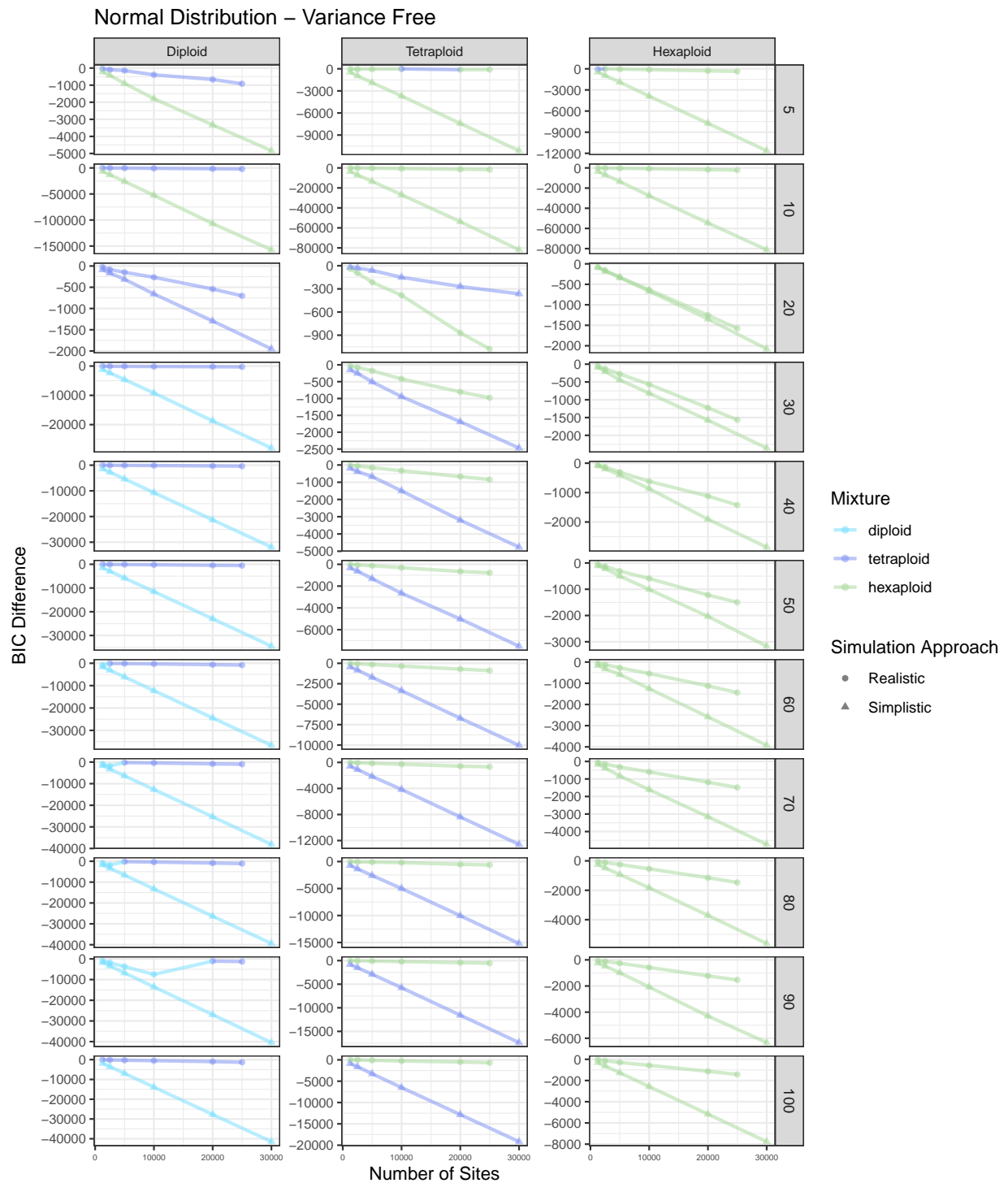


Figure S42. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with variance free.

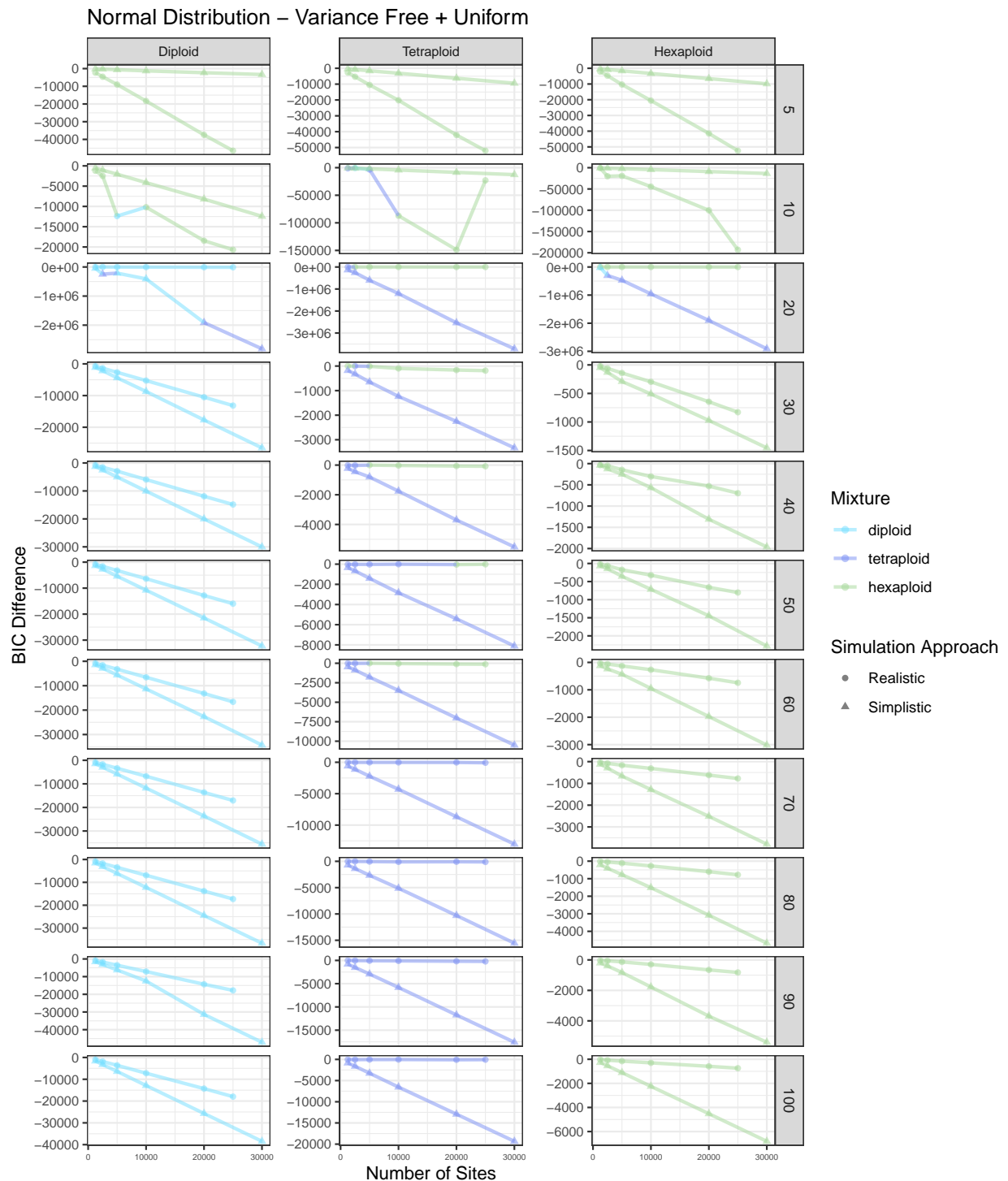


Figure S43. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a normal distribution with variance free with a uniform mixture.

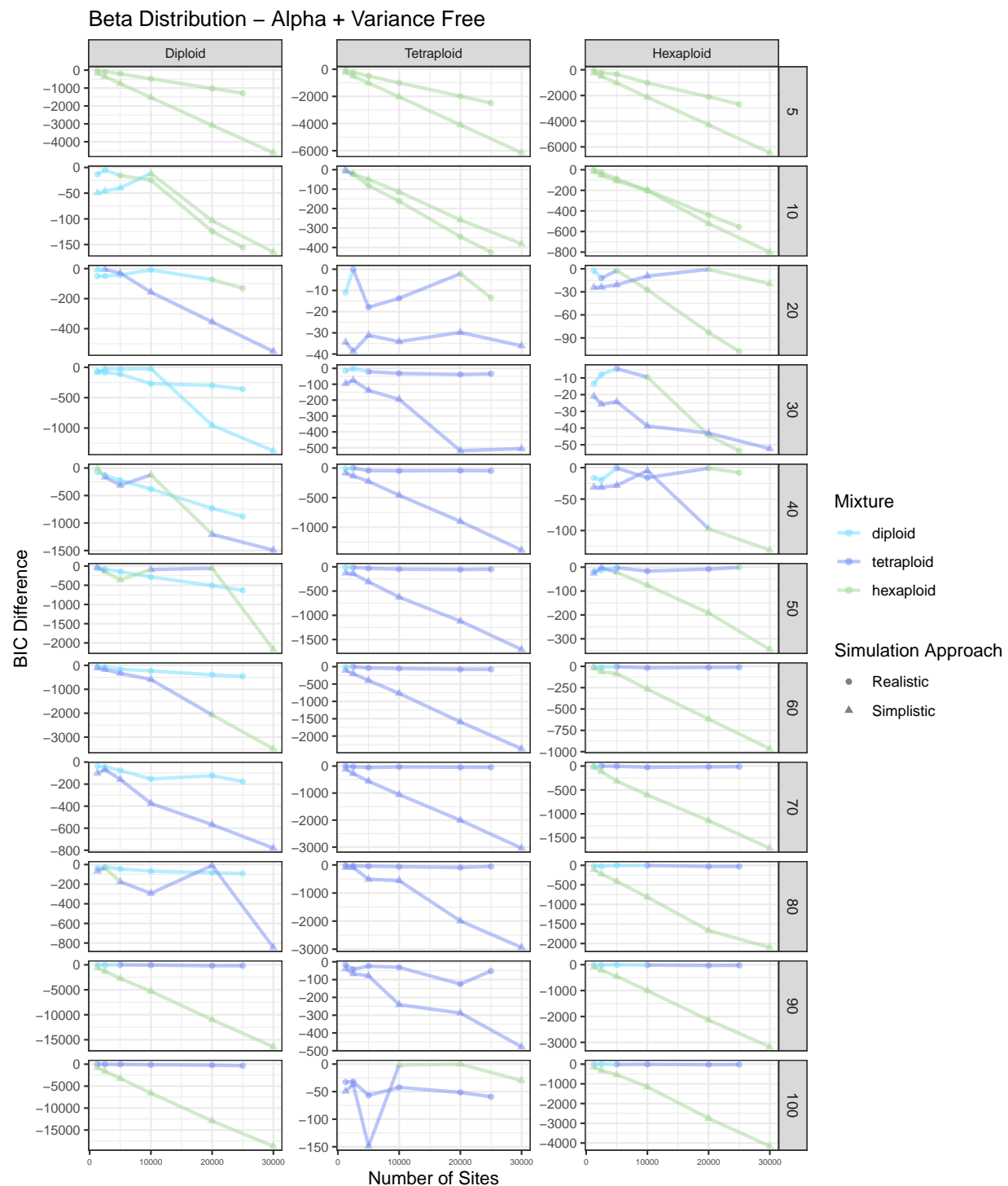


Figure S44. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha and variance free.

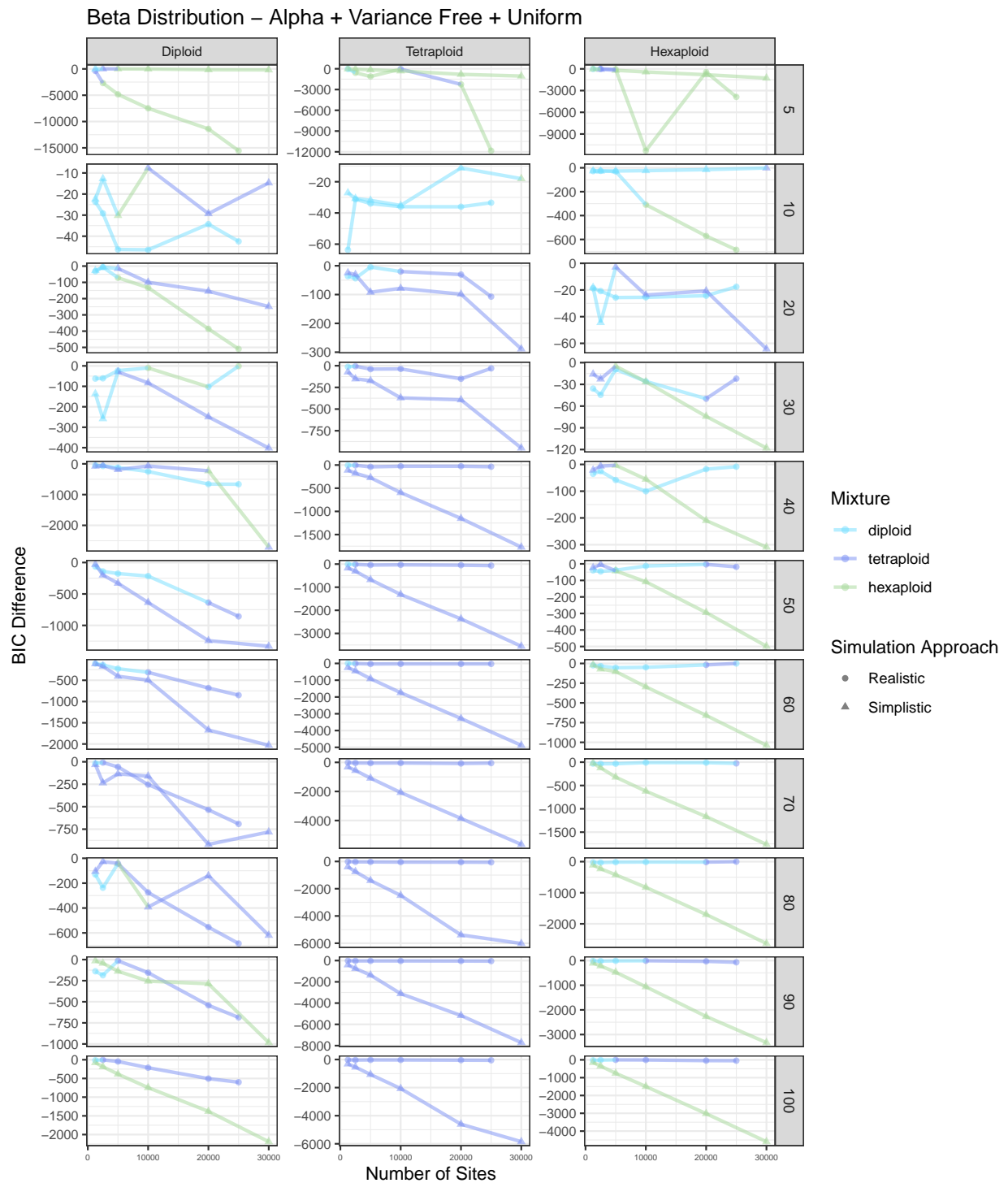


Figure S45. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha and variance free with a uniform mixture.

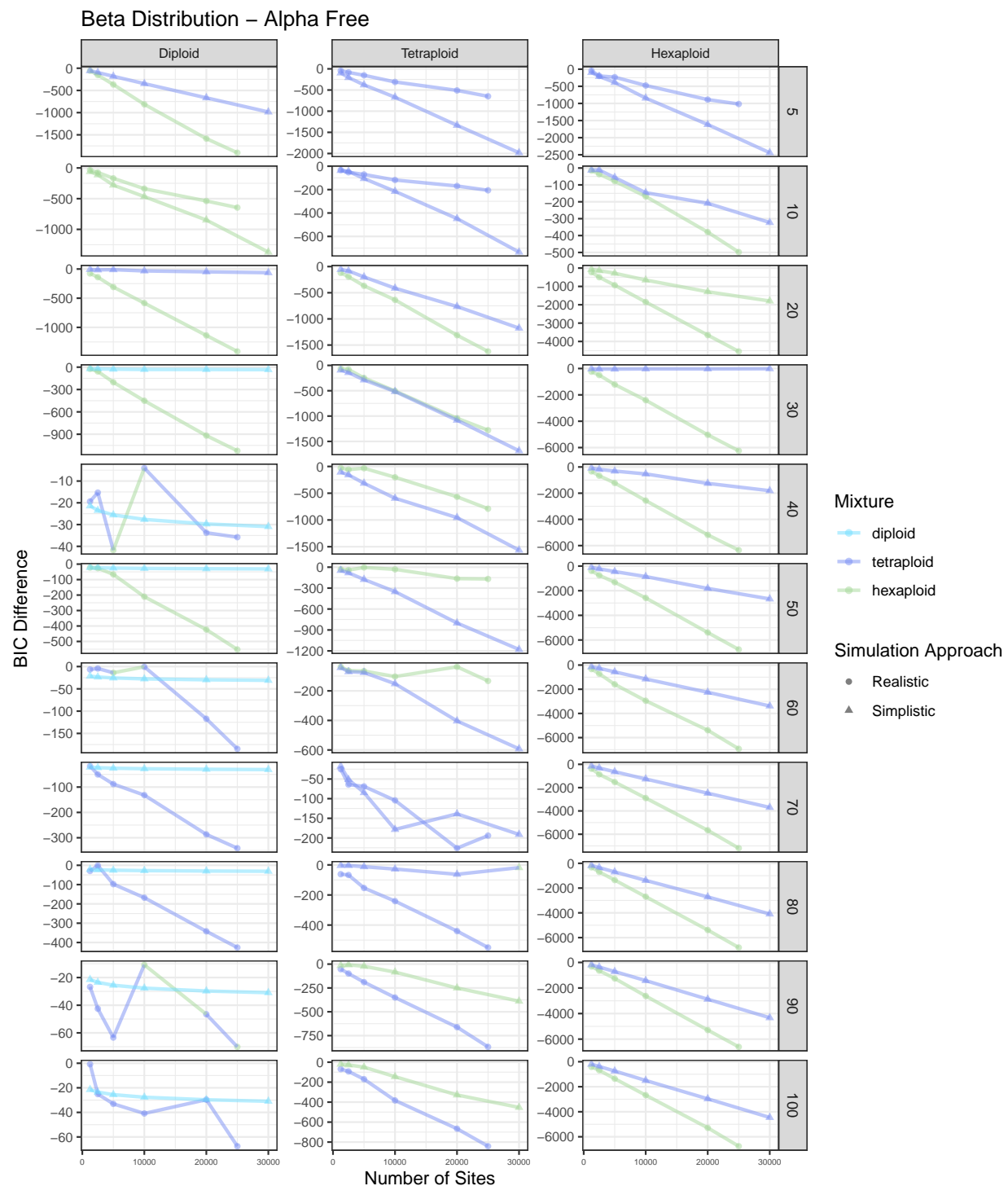


Figure S46. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha free.

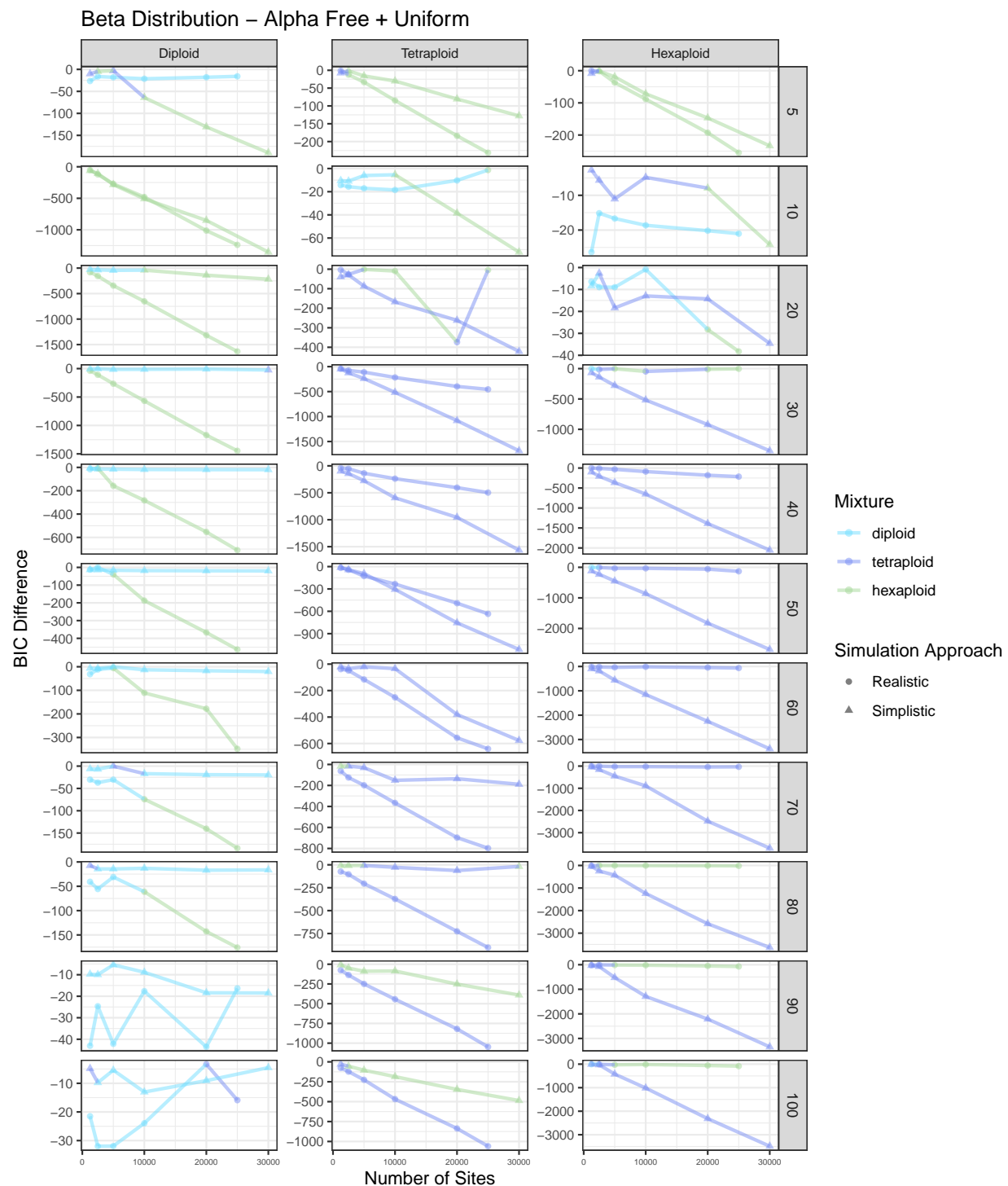


Figure S47. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with alpha free with a uniform mixture.



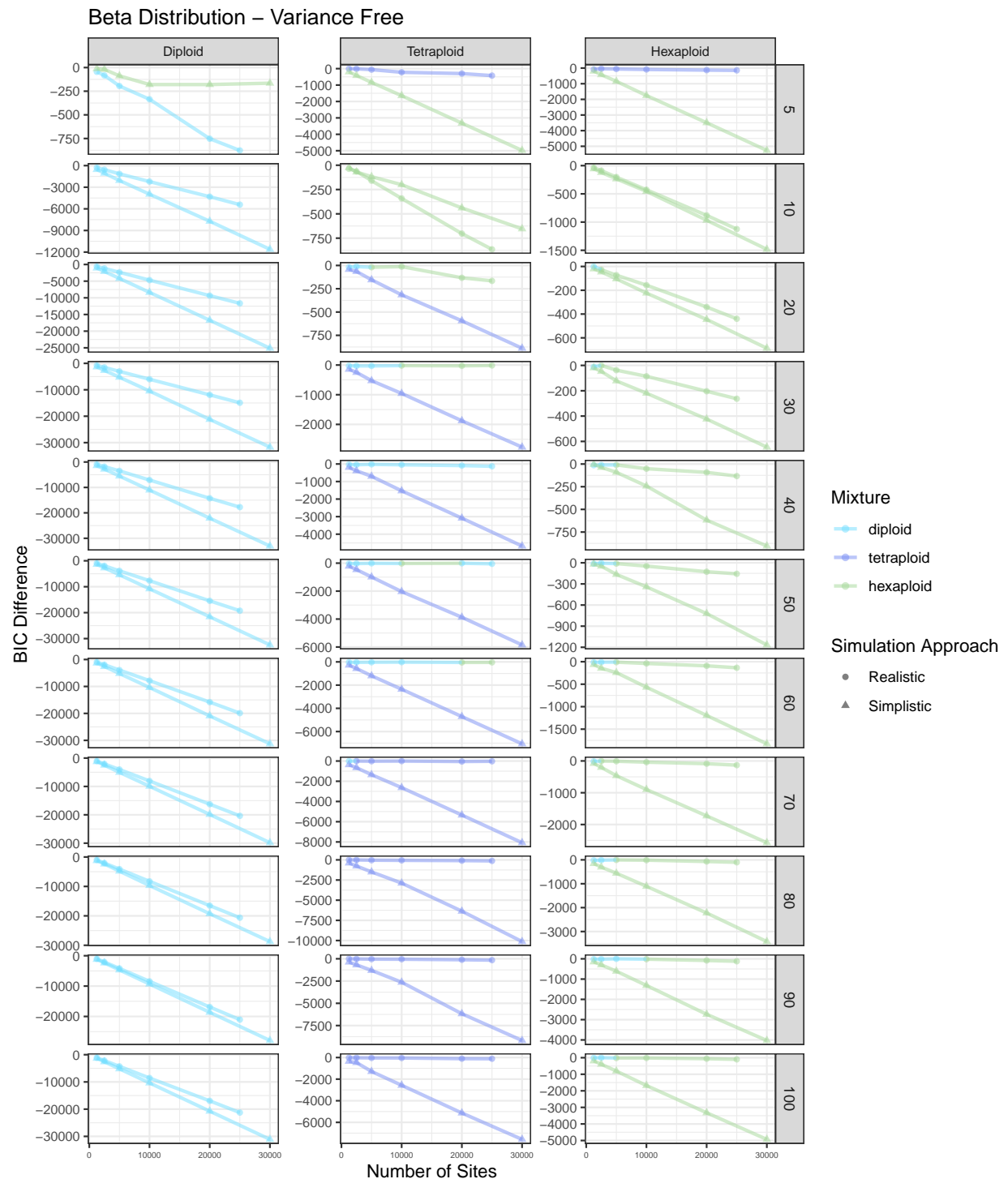


Figure S48. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with variance free.

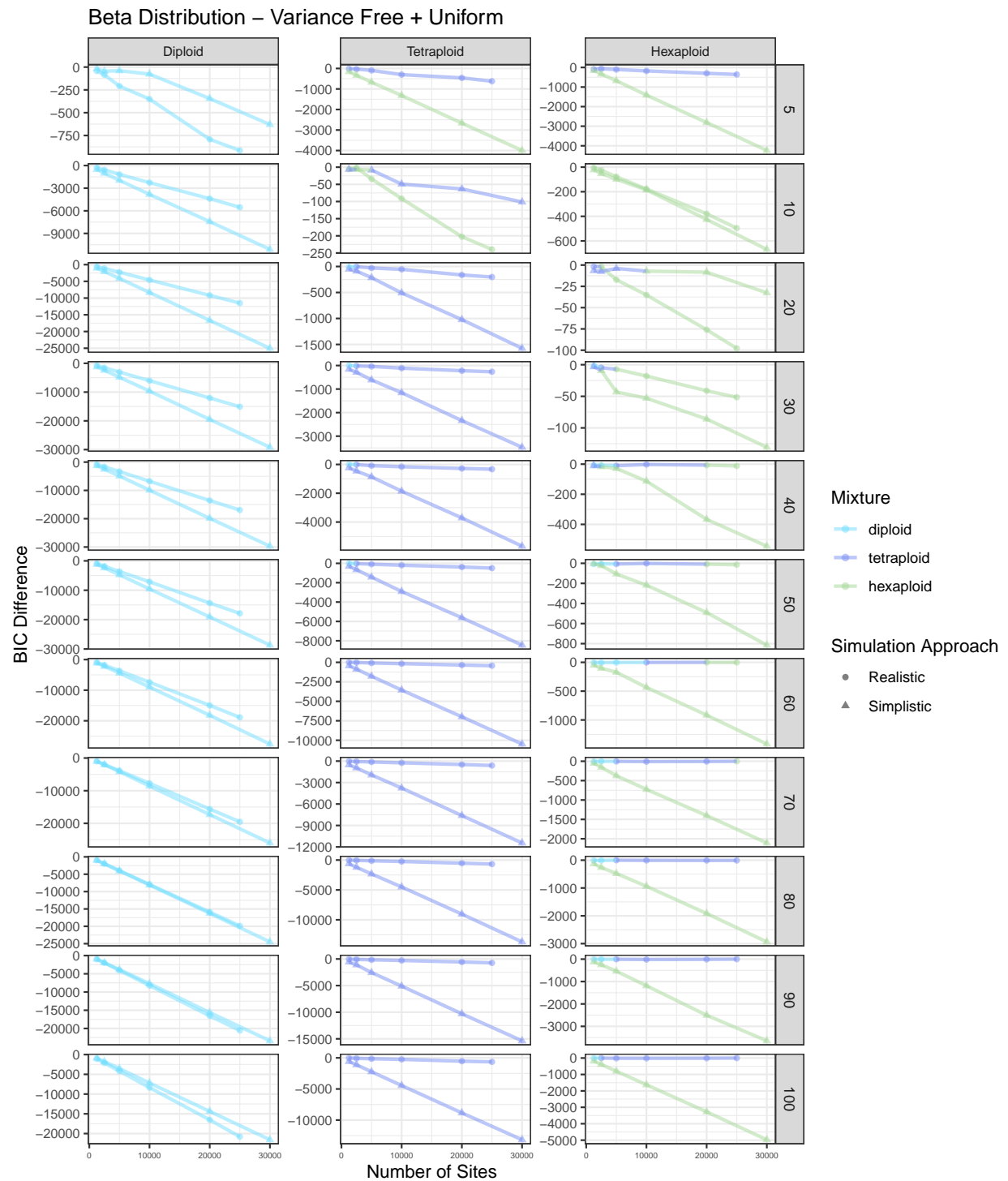


Figure S49. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta distribution with variance free with a uniform mixture.

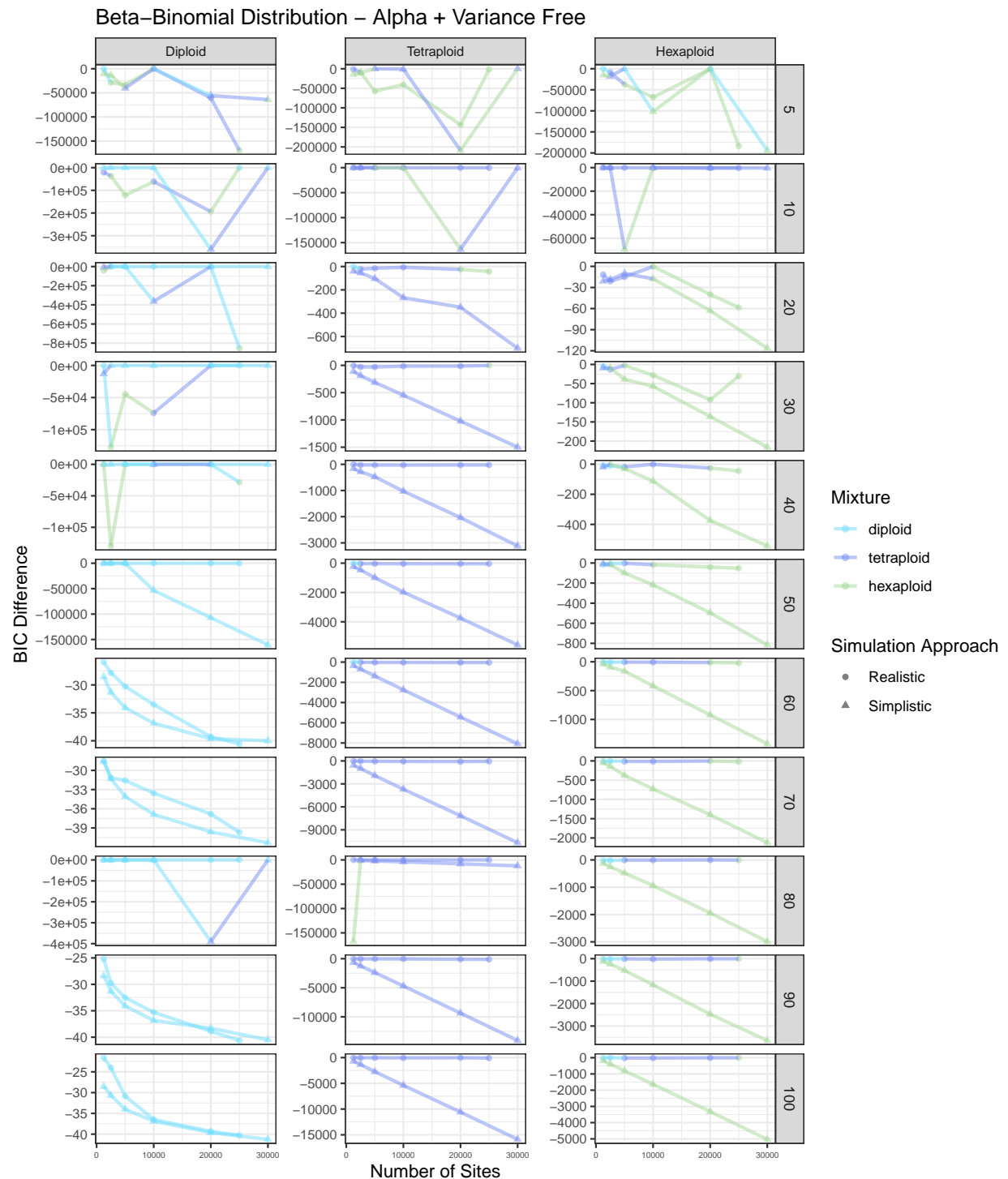


Figure S50. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha and variance free.

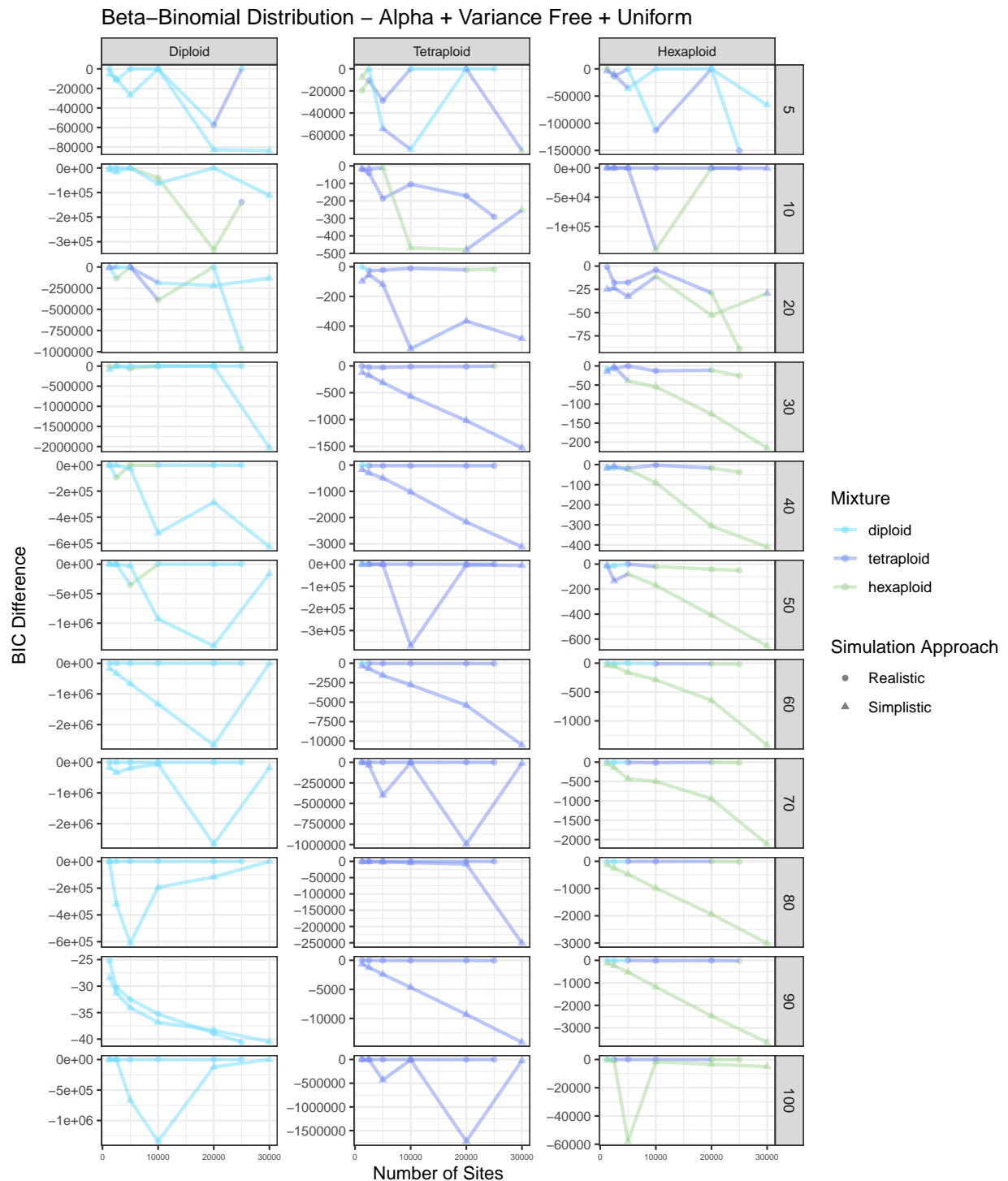


Figure S51. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha and variance free with a uniform mixture.

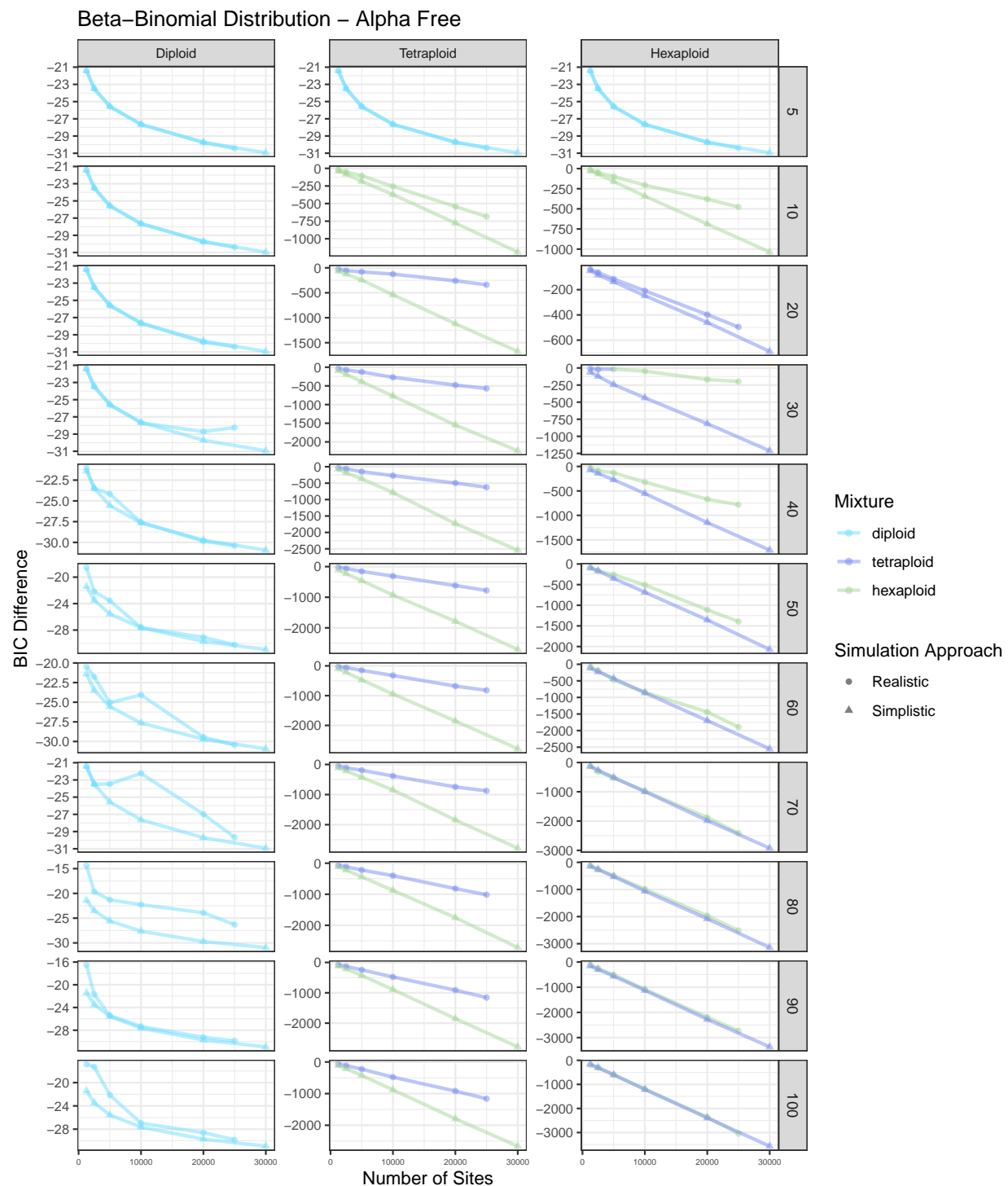


Figure S52. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha free.

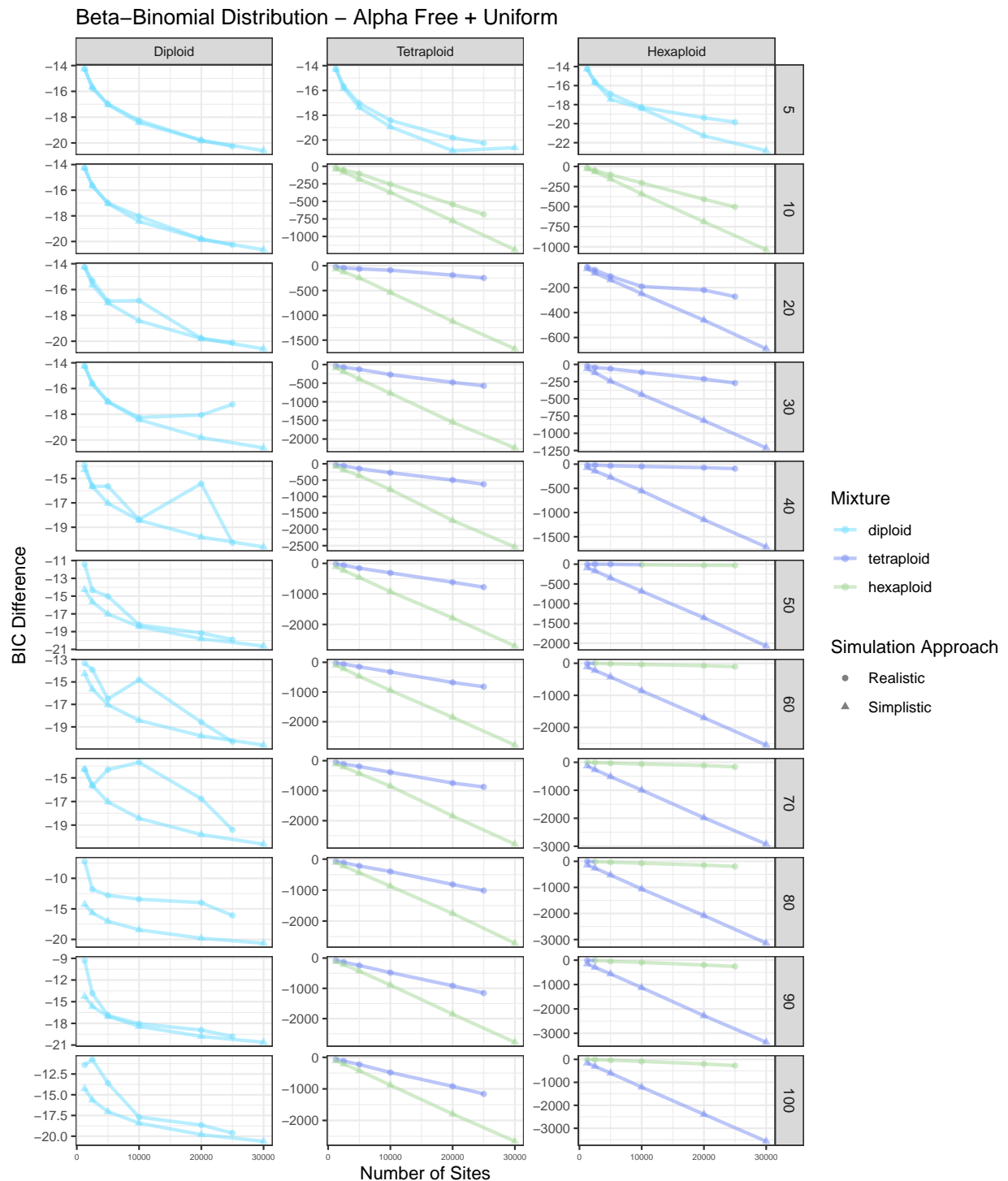


Figure S53. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with alpha free with a uniform mixture.

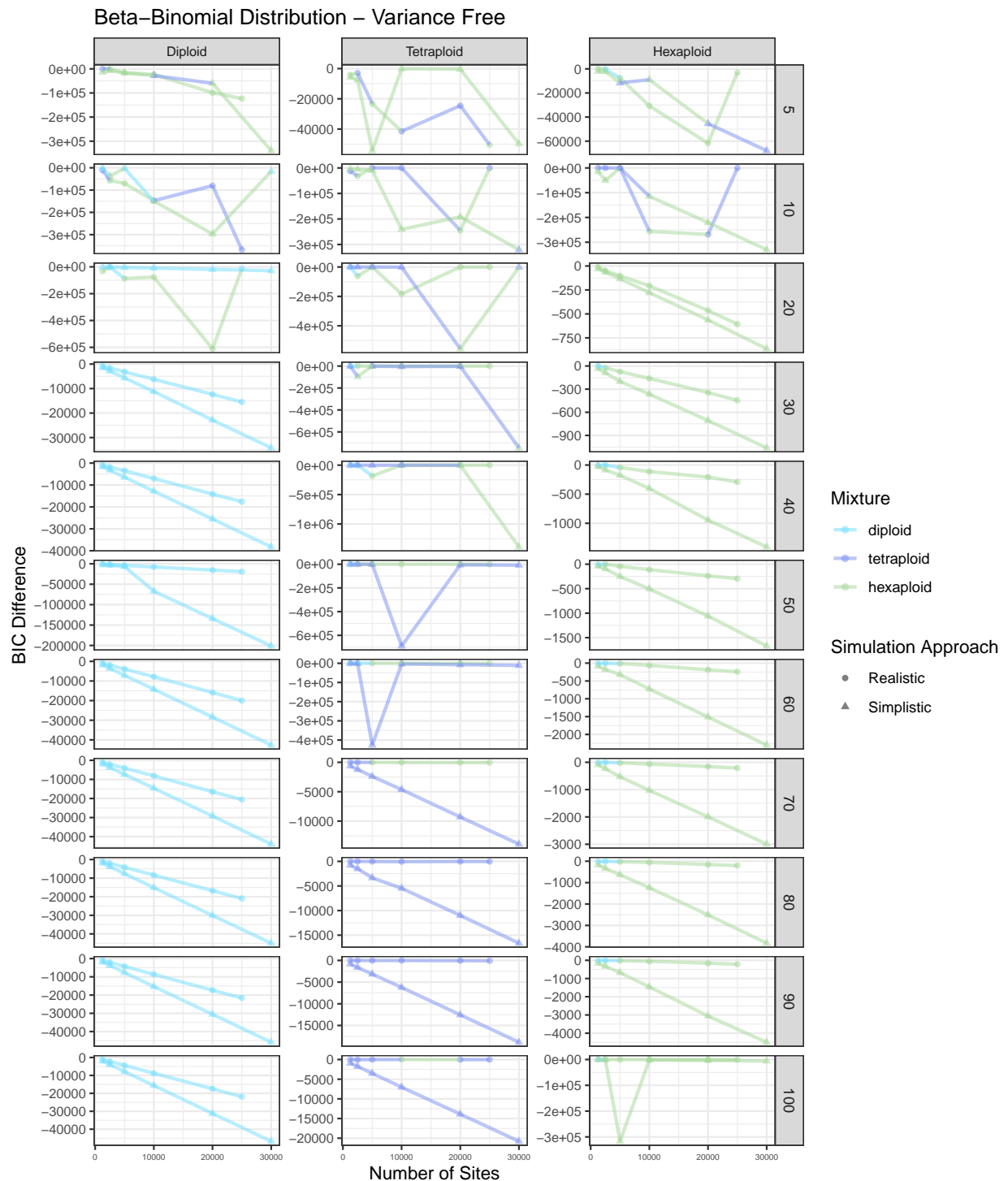


Figure S54. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with variance free.

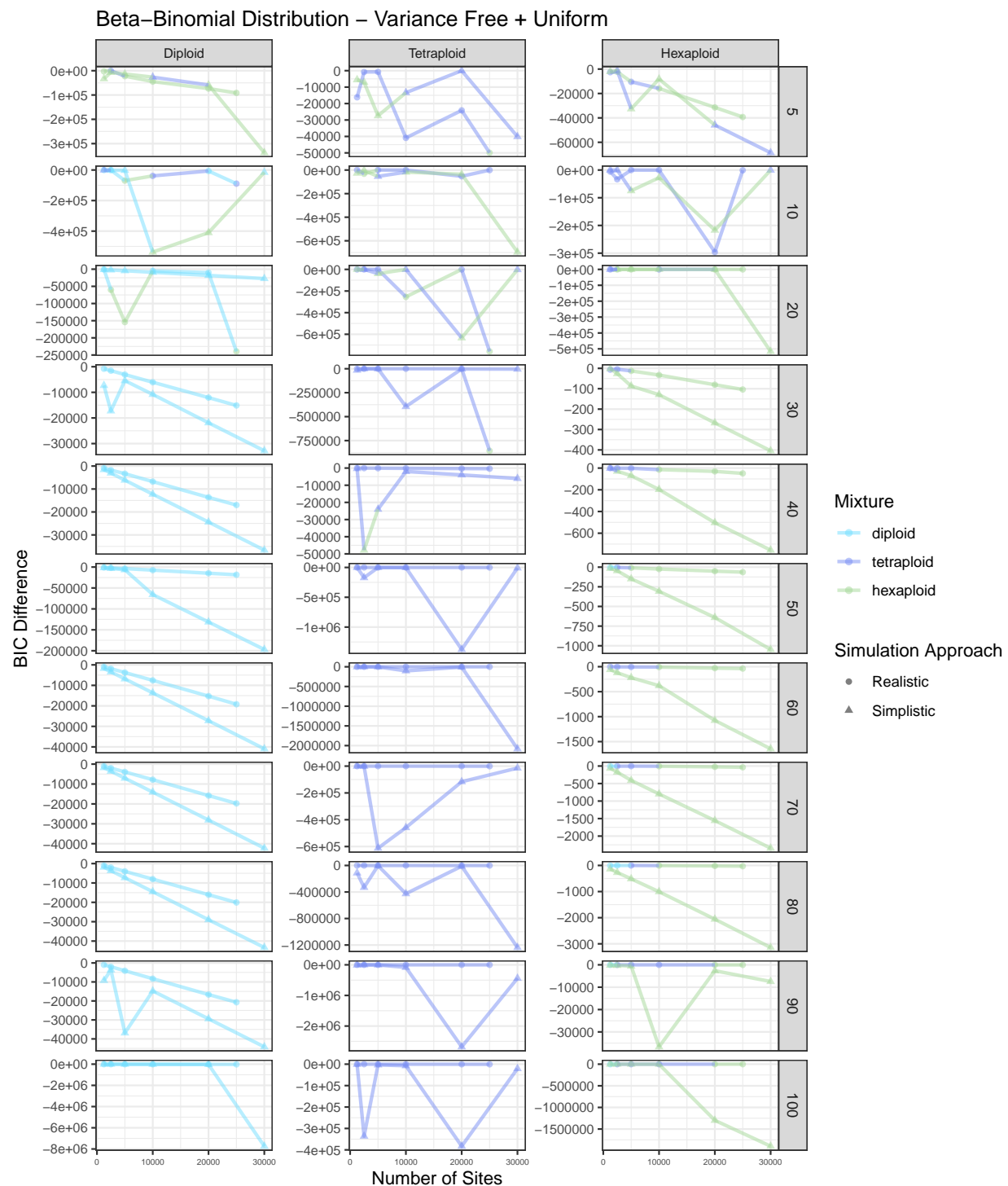


Figure S55. BIC score difference between the best and second best model for simulated diploid, tetraploid, and hexaploid samples across different numbers of sites for eleven different coverage amounts. The color of each point represents the best model. The shape of each point represents the approach used to simulate that sample. This represents a beta-binomial distribution with variance free with a uniform mixture.



Figure S56–S58. — Probability of the correct model choice given the BIC difference between the best and second-best model for all simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples. The probability of success was predicted based on a logistic regression where accuracy is a function of BIC difference. We expect the BIC difference between the best and second-best model to increase with the probability of success when the BIC difference is indicative of the model’s accuracy.

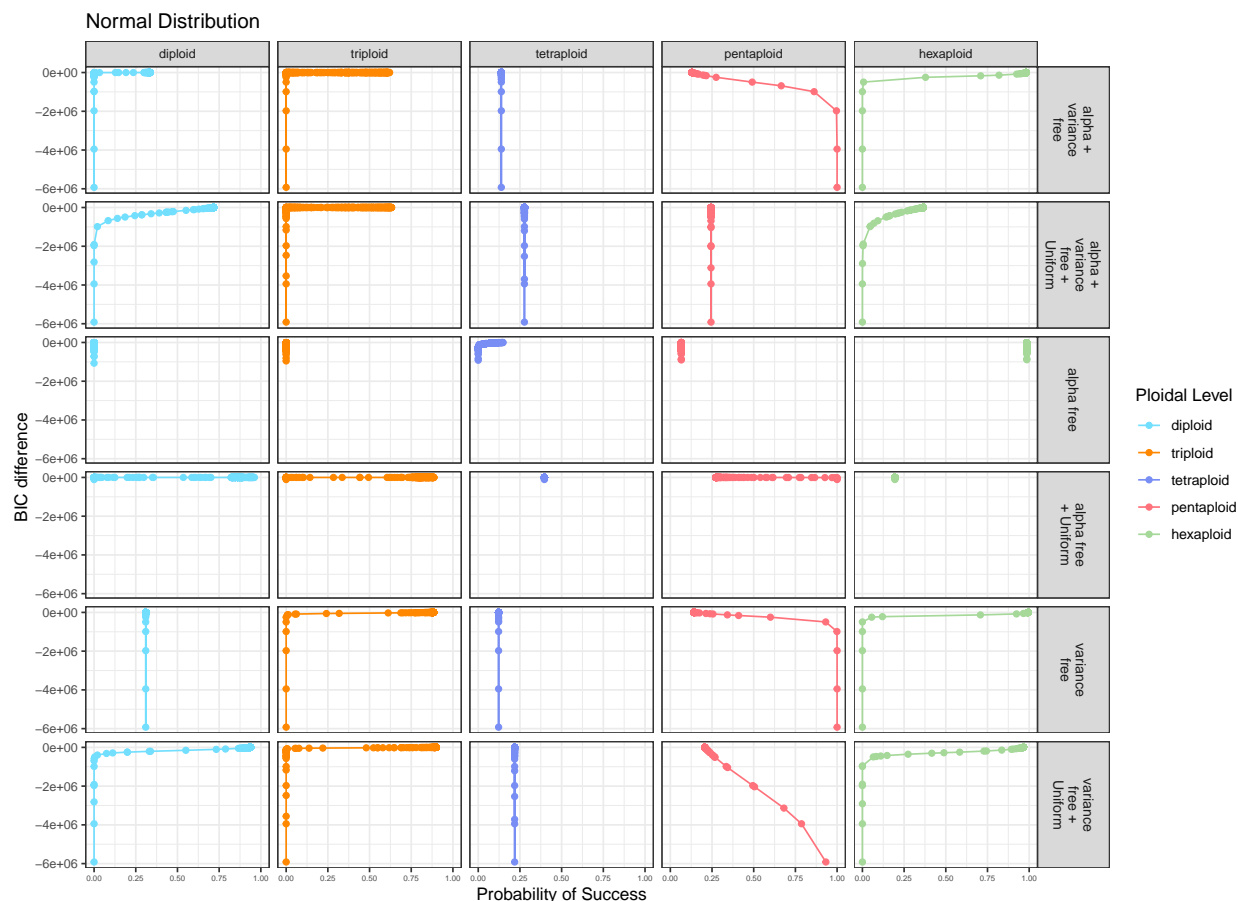


Figure S56. Probability of the correct model choice given the BIC difference between the best and second best model for all simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples. The probability of success was predicted based on a logistic regression where accuracy is a function of BIC difference.

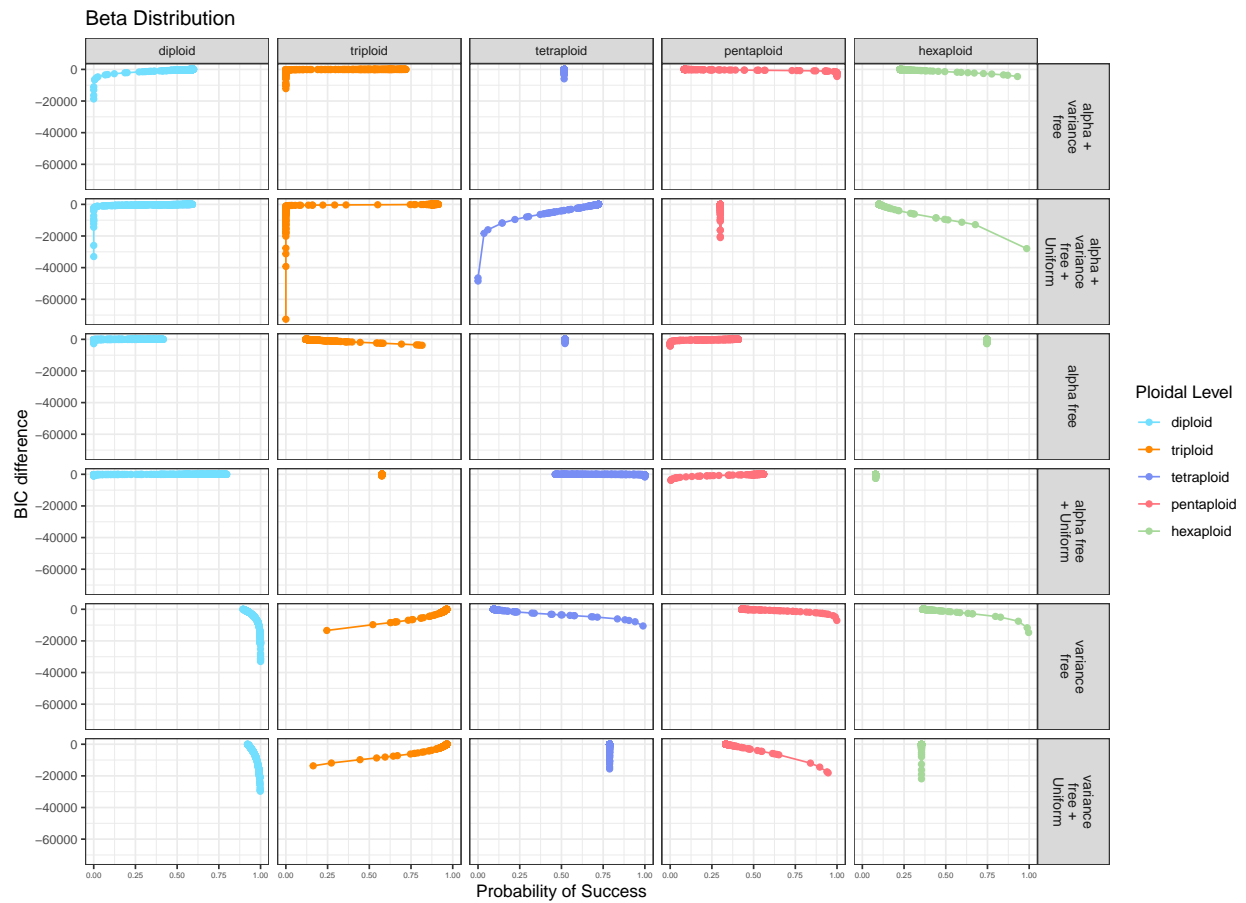


Figure S57. Probability of the correct model choice given the BIC difference between the best and second best model for all simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples. The probability of success was predicted based on a logistic regression where accuracy is a function of BIC difference.

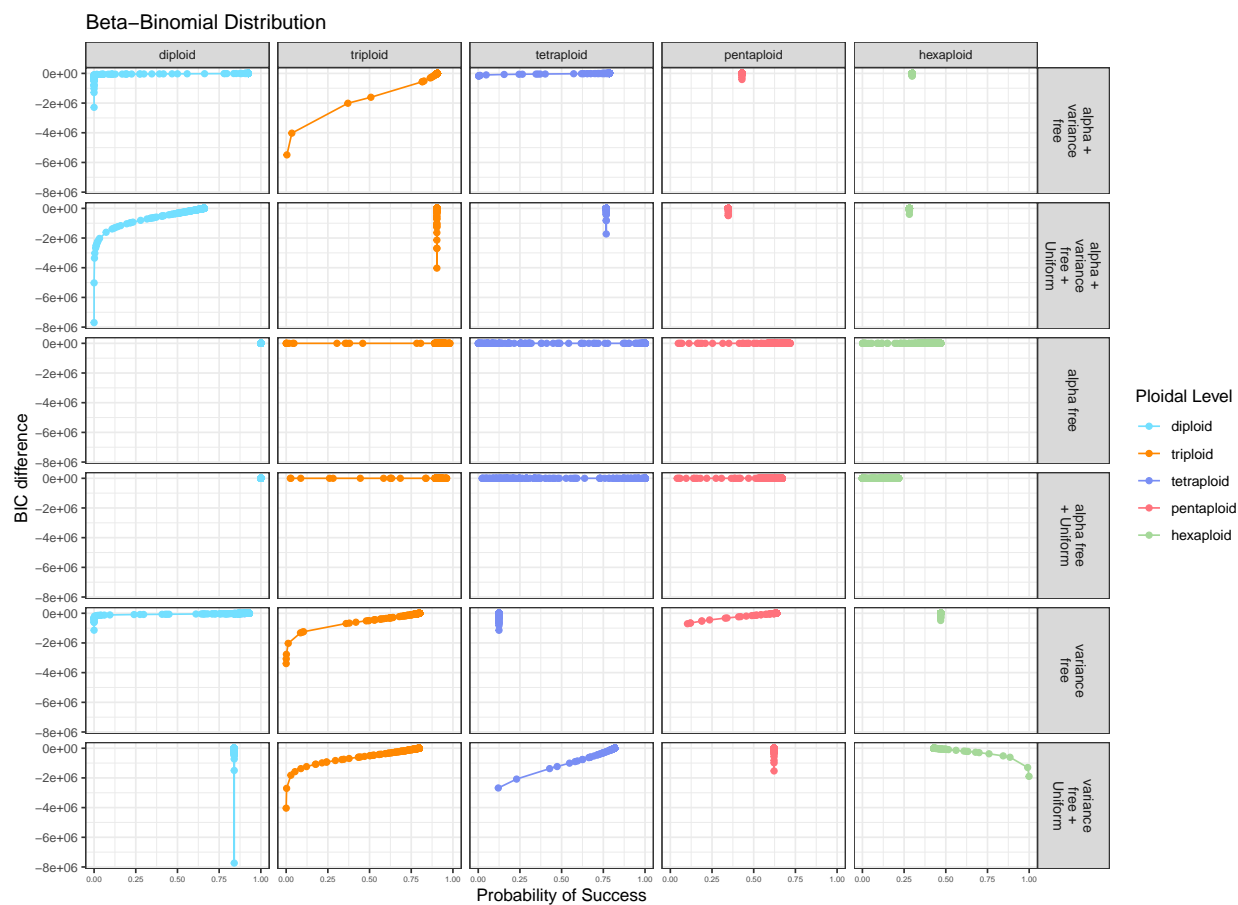


Figure S58. Probability of the correct model choice given the BIC difference between the best and second best model for all simulated diploid, triploid, tetraploid, pentaploid, and hexaploid samples. The probability of success was predicted based on a logistic regression where accuracy is a function of BIC difference.

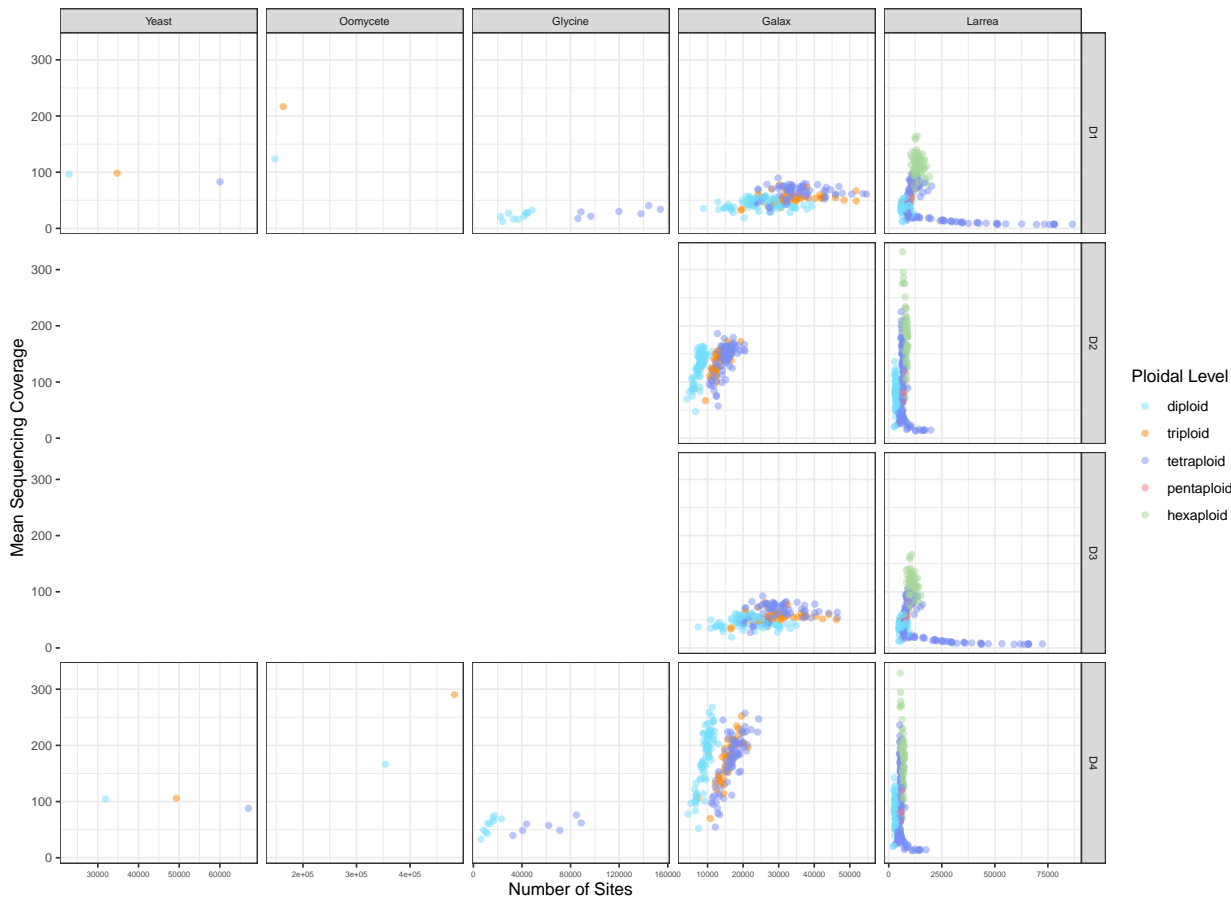


Figure S59. Number of sites and mean sequencing coverage for all data sets given the filtering type (see in-text for description).