SUPPLEMENTARY NOTE

FRAMEWORK OF *CHROMO.CRAWL*

Chromo.Crawl is one of the primary functions of *PhyloWGA* that seeks to test for the presence of "supergenes" (i.e., contiguous loci with evidence of a shared genealogical history) across whole-genome alignments (WGAs). In this section, we discuss several major benefits of the *Chromo.Crawl* framework when compared to standard "statistical binning" procedures (i.e., Mirarab *et al*. 2014; Bayzid *et al*. 2015; Liu and Edwards 2015; Adams and Castoe 2019a) that have been developed for a similar purpose. *Chromo.Crawl* acts by "crawling" across a WGA from one genomic window to the next adjacent one, and for each pair of windows, a likelihood ratio test is applied using CONCATEPILLAR (Leigh *et al.*, 2008). These two features of *Chromo.Crawl* (crawling across contiguous genomic windows and implementing a likelihood ratio test) represent important and distinct advantages over standard statistical binning procedures, such as the pipelines described in (Mirarab *et al*. 2014; Bayzid *et al*. 2015; Liu and Edwards 2015; Adams and Castoe 2019a,b).

First, the procedure of crawling from one genomic region to the next means that the specific genomic location and context are considered by *Chromo.Crawl*. Recombination acts to effectively decouple the genealogical histories of adjacent loci along a chromosome, such that the genealogical histories of loci separated by a recombination event will be unlinked. Conversely, two loci will share the same underlying gene tree in the absence of recombination. Because the probability of crossing over between two loci is a function of their distance to one another, adjacent loci are more likely to share a common genealogical history than more distant loci, and the "crawling" behavior of *Chromo.Crawl* is effectively designed to incorporate this spatial information while testing for phylogenetic congruency along a WGA. This procedure is in stark contrast to typical techniques for statistical binning that use

differences in estimated topologies among gene trees obtained across distant genomic regions and even chromosomes (Bayzid *et al.*, 2015; Adams and Castoe, 2019). Indeed, the recent studies of Adams and Castoe (2019a,b) found evidence that standard statistical binning is likely to bias species tree estimates by incorrectly constructing supergenes from loci sampled from distant genomic regions and diverse chromosomes.

Another primary advantage of the *Chromo.Crawl* framework when compared to traditional binning techniques is the use of a formal likelihood ratio test for assessing evidence of phylogenetic congruency among loci. CONCATEPILLAR is based on the principles of likelihood ratio tests of topological congruency, such as Huelsenbeck and Bull (1996), and therefore benefits from the statistical foundation provided by such tests, lending itself to predictable statistical properties. In comparison, standard statistical binning pipelines typically use an *ad hoc* procedure based on differences in topologies among bootstrap replicates. For standard statistical binning, if the fraction of bootstrap replicates leading to phylogenetic incongruence is below a certain user-defined cutoff, then the loci are deemed to be congruent. *Chromo.Crawl* instead uses CONCATEPILLAR to test whether a model of congruency (i.e., single genealogical history) or incongruency (i.e., distinct genealogical histories) is a better fit for a given alignment. If the data support a congruent model for adjacent genomic windows, then the set of contiguous loci are concatenated together as a supergene, and conversely, the loci are deemed independent if a model of discordance is supported based on the likelihood ratio.

BENCHMARKING

The computational complexity of the functions of *PhyloWGA* is highly dynamic and depends upon a number of factors, including (but not limited to): (1) number of taxa in the WGA, (2) total length of the WGA, (3) experimental design (i.e., number, length, and spatial distribution of genomic windows),

(4) evolutionary model complexity (e.g., JC69 versus GTR model versus model selection, and recombination rate), and (5) particular analysis type (*Chromo.Phylome* versus *Chromo.Crawl*). Like most standard phylogenetic analyses, the complexities of the algorithms scale with both the number of taxa and the total length of the WGA, whereas the particular experimental design (i.e., distribution of windows defined by the user) may have dynamic effects on computational speed. For example, *Chromo.Phylome* using a single, long window is likely to be faster than if that window is partitioned into multiple shorter windows because *Chromo.Phylome* will be applied to each window separately (i.e., phylogenetic tree model will be fit to each window instead of a single window). The particular evolutionary model used will also impact the efficiency of *PhyloWGA* by modulating the number of parameters estimated for each window. *Chromo.Phylome* conducts phylogenetic inference across genomic windows, whereas *Chromo.Crawl* applies tests of congruency among windows. Thus, the *Chromo.Crawl* function will run substantially slower than *Chromo.Phylome* because it both infers trees and implements a likelihood ratio test, and thus, the efficiency of *Chromo.Crawl* will also likely fluctuate as it crawls along a chromosome. For example, we ran two versions of *Chromo.Phylome* on the primate WGA using a single thread of a 2.8 GHz CPU that represent two different nucleotide model settings, and we found the following run times: GTR model for all windows (run time of 10.35 hours) and GTR+Γ model for all windows (run time of 22.07 hours). That is, estimating the shape parameter of the Γ model of among-site variation approximately doubled the running time for this example. Additionally, we found that our *Chromo.Crawl* simulation demonstration (described in the next section and results shown in Figures S1-S4) of 100 kb alignments ran for an average of approximately 2.4 hours each using a 2.8 GHz CPU with two threads. The *Chromo.Phylome* analyses for these 100 kb simulated WGAs were far quicker (i.e., each ran under a minute using a single thread of a 1.7 GHz Dual-Core Intel Core i7 CPU).

In any case, the functions *Chromo.Phylome* and *Chromo.Crawl* both provide indications of computational time required for analyses. For example, *Chromo.Phylome* provides an estimate of the total time needed for analysis (based on the first window), and the percentage of total windows that have been analyzed as the algorithm proceeds along the WGA. Similarly, *Chromo.Crawl* tracks the progress of the algorithm by printing the percentage of total windows that have been "crawled" over, and *Chromo.Crawl* also allows the user to specify the number of cores with the "numeric.NumberOfCores" argument. Finally, *PhyloWGA* now includes a function *Organize.ParallelPhyloWGA* that streamlines and organizes WGAs and *PhyloWGA* scripts for parallel analyses. This function partitions a WGA into a number of user-defined subsets placed within directories for easy execution and parallel analysis.

EXPLORING THE ACCURACY OF *PHYLOWGA* ON SIMULATED WGAS

To explore the accuracy of *PhyloWGA*, we conducted an array of simulation analyses that are inspired by the Primate dataset (shown in Figure 1), and that varied in recombination rate *r*. We simulated genealogies along 100 kb alignments with the program ms (Hudson, 2002) using a 10-taxon species tree inspired by the relationships of Human, Chimpanzee, Gorilla, Orangutan, Macaque, Marmoset, Tarsier, Galago, Lemur, and Rat provided in a previous study (Song, *et al.,* 2012) and three different recombination rate values (r = 10^{-9}, 10^{-8}, or 10^{-7} per site per generation). For each genealogy and associated alignment block output by ms (i.e., subsets of the 100 kb alignment separated by recombination events), we then simulated nucleotide sequence alignments using the program Seq-Gen (Rambaut and Grass, 1997) and a HKY model (Hasegawa *et al.*, 1985) with the following parameters: transition/transversion ratio of 4.6, and base equilibrium frequencies of 0.3 (A), 0.2 (C), 0.2 (G), and 0.3 (T). In our simulations, we used the population-scaled mutation rate $\theta = 4N\mu = 0.00104$ for a diploid effective population size N =10^4 (Takahata, 1993) and a mutation rate $\mu = 2.6 \times 10^{-8}$ per site

per generation (Narasimhan *et al.*, 2017). These HKY parameters were inspired by previous studies of primate relationships and species tree analyses (Burgess and Yang, 2008; Koch and DeGiorgio, 2020). These resulting simulated alignments were then concatenated together to form a single, 100 kb WGA fasta file. We next conducted two *PhyloWGA* analyses: (1) *Chromo.Crawl* with one kb windows and step sizes, and (2) using these *Chromo.Crawl* coordinates of concatenated windows to construct gene trees using *Chromo.Phylome* with nucleotide substitution model selection. For each combination of simulation parameters, we repeated the process nine times, and we plotted Robinson-Foulds (RF) distances (Robinson and Foulds, 1981) for each nucleotide site between the true simulated genealogies and their corresponding inferred trees from *Chromo.Phylome*. Additionally, we plotted the location of recombination events that resulted in topology swaps (i.e., red lines and dots in Figures S2-S4) and the location of breakpoints reconstructed with *Chromo.Crawl* (i.e., alternating light and dark gray blocks that represent concatenated windows in Figures S2-S4).

As demonstrated with the simulations, the accuracy of *PhyloWGA* is dynamic in response to the recombination rate *r*. For example, the mean RF distance is much lower under the low recombination rate ($r = 10^{-9}$ per site per generation) simulations when compared with the high recombination rate ($r = 10^{-7}$ per site per generation) scenarios (Figures S1a versus S1c). These results can be observed when comparing the true, simulated recombination breakpoints that result in topology changes (i.e., red lines in Figures S2-S4) with the inferred breakpoints recovered by *PhyloWGA* (alternating light and dark gray blocks in Figures S2-S4). For example, the inferred and true recombination breakpoints appear to be more accurately reconstructed with the low recombination simulations (Figure S2) compared to the high recombination rate simulations (Figure S4). Under the large recombination rate ($r = 10^{-7}$ per site per generation), the lengths of recombination-free nucleotide stretches are small because there are a large number of observed recombination events across the alignment (Figure S4).

A fundamental goal of *Chromo.Crawl*, and *PhyloWGA* more generally, is to improve the accuracy of chromosome-scale phylogenetic analyses by flexibly and adaptively incorporating information about gene tree signal and variability of genomic regions. Thus, we sought to understand its performance at achieving this goal by comparing its accuracy with three alternative strategies that do not consider evidence for (or against) shared gene trees among adjacent windows: (1) trees inferred across an alignment using a fixed window size ("FIXED"), (2) trees inferred across random windows without regards to location ("RANDOM"), and (3) a single tree is inferred by concatenating the entire 100kb alignment ("CONCAT"), such that gene tree variability is ignored. The RANDOM approach can be described in three steps: (1) randomly sample a starting site (with uniform probability across the 100kb alignment) to denote beginning of a window, (2) define the end of a window by the site position that is located one, two, four, five, 10, or 20 kb downstream of the start site sampled in the previous step, and (3) this process is repeated 25 times to obtain 25 windows that are positioned randomly throughout the chromosome. *Chromo.Crawl* represents an adaptive, genome-informed approach designed to respond to shared phylogenetic signal (or lack thereof) among adjacent loci, and thus, we predicted that it would yield more accurate inferences when compared to the either the FIXED or RANDOM procedures, as well as the CONCAT approach that ignores gene tree variation. We simulated nine 100kb replicate datasets according to the same procedures as before (i.e., Figures S1-S4), and we applied each of the four strategies (*Chromo.Crawl,* FIXED, RANDOM, and CONCAT) across a range of different window sizes (one, two, four, five, 10, and 20 kb) for the low ($r = 10^{-9}$), medium ($r = 10^{-8}$), and high recombination rates ($r = 10^{-7}$). We measured the mean RF distance (between the true and inferred tree at each site along the alignment) across replicates to quantify differences between the three methods in terms of phylogenetic accuracy. That is, we sought to understand whether the core function of *Chromo.Crawl* (i.e., infer more accurate gene trees using longer, concatenated windows) was indeed successful at improving phylogenetic accuracy across simulated chromosomes, and in particular, its

success when compared to the RANDOM, FIXED, and CONCAT approaches that are agnostic to diverse spans of phylogenetic signal across an alignment.

Our results provide evidence that *Chromo.Crawl* does indeed provide more accurate inferences than either the RANDOM, FIXED, or CONCAT approaches (Figure S5). For example, the minimum mean RF distance of *Chromo.Crawl* was always smaller than or comparable to the corresponding window size of both the FIXED and RANDOM in the majority of cases. This is perhaps best illustrated when comparing the one kb results of *Chromo.Crawl* (mean RF = 0.09) to the one kb FIXED (mean RF = 0.37) and RANDOM (mean RF = 0.38) analyses, respectively, for the low recombination results (Figure S5a). Accuracy tends to increase for both the FIXED and RANDOM strategies with larger window sizes (i.e., approximately 10 to 20 kb), while in contrast, *Chromo.Crawl* appears to adaptively adjust to the optimal window size across each recombination rate setting, finding a minimum mean RF distance with windows of size two kb for the low recombination rate (Figure S5a), one kb for the medium recombination rate (Figure S5b), and four to 20 kb for the high recombination rate (Figure S5c). Moreover, even the smallest minimum window size for *Chromo.Crawl* of one kb performed comparatively well, indicating that it would be sufficient for a user to just specify a minimum window size of one kb, with *PhyloWGA* adaptively changing window sizes across a chromosome and providing improved phylogenetic accuracy over fixed user-defined window approaches. The pervasive abundance of recombination events (e.g., see Figure S4 as an illustration) reduced phylogenetic accuracy across the board in our highest recombination rate simulations (Figure S5c). Yet, we found evidence that *Chromo.Crawl* nonetheless outperformed the FIXED, RANDOM, and CONCAT approaches in these challenging scenarios. These results suggest that the genome-informed approach of *Chromo.Crawl* does indeed provide meaningful improvements in phylogenetic accuracy over alternative approaches that do not attempt account for the tendency of adjacent genomic regions to share a common genealogical history.

Collectively, these results (Figures S1-S5) promote the user-friendly framework *PhyloWGA* as a tool for improving chromosome-scale phylogenetic accuracy. In general, we expect the accuracy of *PhyloWGA* to be dynamic as a function of both experimental and evolutionary parameters (i.e., Figs. S2-S5). As expected, these results suggest that genomic regions with low recombination rates may be more accurately reconstructed with *PhyloWGA* (and most any other approach, including RANDOM and FIXED strategies; Figure S5), when compared with regions with high recombination rate. In any case, we encourage users to carefully consider the evolutionary context (e.g., recombination rates) of their particular datasets when analyzing with *PhyloWGA* or any other phylogenetic analyses.

REFERENCES

Adams,Richard H. and Castoe,T.A. (2019) Statistical binning leads to profound model violation due to gene tree error incurred by trying to avoid gene tree error. *Mol. Phylogenet. Evol.*, **134**, 164-171.

Adams,Richard H and Castoe,T.A. (2019) Supergene validation: A model-based protocol for assessing the accuracy of non-model-based supergene methods. *MethodsX*, **6**, 2181–2188.

Bayzid,M.S. *et al.* (2015) Weighted statistical binning: enabling statistically consistent genome-scale phylogenetic analyses. *PLoS One*, **10**, e0129183.

Burgess,R. and Yang,Z. (2008) Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.*, **25**, 1979–1994.

Hasegawa,M. *et al.* (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, **22**, 160–174.

183 Hudson,R.R. (2002) Generating samples under a Wright–Fisher neutral model of genetic variation.
184 *Bioinformatics*, **18**, 337–338.

185 Huelsenbeck,J.P. and Bull,J.J. (1996) A likelihood ratio test to detect conflicting phylogenetic signal.
186 *Syst. Biol.*, **45**, 92–98.

187 Koch,H. and DeGiorgio,M. (2020) Maximum likelihood estimation of species trees from gene trees in
188 the presence of ancestral population structure. *Genome Biol. Evol.*, **12**, 3977–3995.

189 Leigh,J.W. *et al.* (2008) Testing congruence in phylogenomic analysis. *Syst. Biol.*, **57**, 104–115.

190 Liu,L. and Edwards,S. V (2015) Comment on "Statistical binning enables an accurate coalescent-based
191 estimation of the avian tree". *Science*, **350**, 171.

192 Mirarab,S. *et al.* (2014) Statistical binning enables an accurate coalescent-based estimation of the avian
193 tree. *Science*, **346**, 1250463.

194 Narasimhan, V. M., *et al.* (2017). Estimating the human mutation rate from autozygous segments
195 reveals population differences in human mutational processes. *Nat. com.*, ***8***, 1–7.

196 Rambaut,A. and Grass,N.C. (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA
197 sequence evolution along phylogenetic trees. *Bioinformatics*, **13**, 235–238.

198 Robinson,D.F. and Foulds,L.R. (1981) Comparison of phylogenetic trees. *Math. Biosci.*, **53**, 131–147.

199 Song, Sen, Liang Liu, Scott V. Edwards, and Shaoyuan Wu. (2012) "Resolving conflict in eutherian
200 mammal phylogeny using phylogenomics and the multispecies coalescent model." *Proc. Natl.*
201 *Acad. Sci.*, **37**, 14942–14947.

202 Takahata, Naoyuki. (1993) "Allelic genealogy and human evolution." *Mol. Biol. Evol.,* **10**: 2–22.

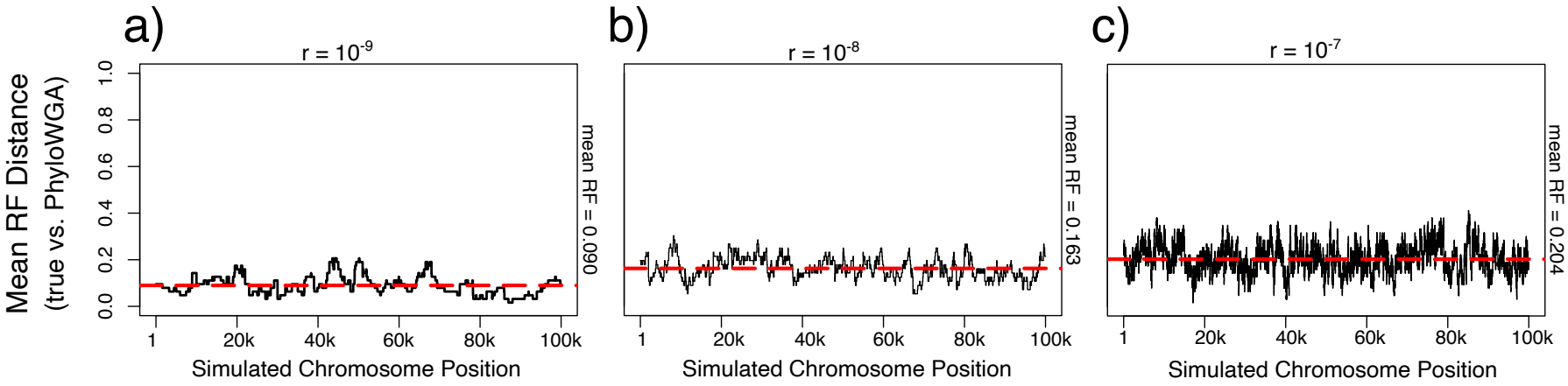

Figure S1. Results of simulation analyses for demonstrating the accuracy of *PhyloWGA* on chromosome alignments simulated with a recombination rate of $r = 10^{-9}$ (a), 10^{-8} (b), and 10^{-7} (c) per site per generation. Black lines indicate the mean Robinson-Foulds (RF) distance for each nucleotide site position in the 100 kb simulated alignments (i.e., site means measured across the respective nine replicates shown in Figures S2-S4). Red dashed line indicates the total mean across the entire alignment.

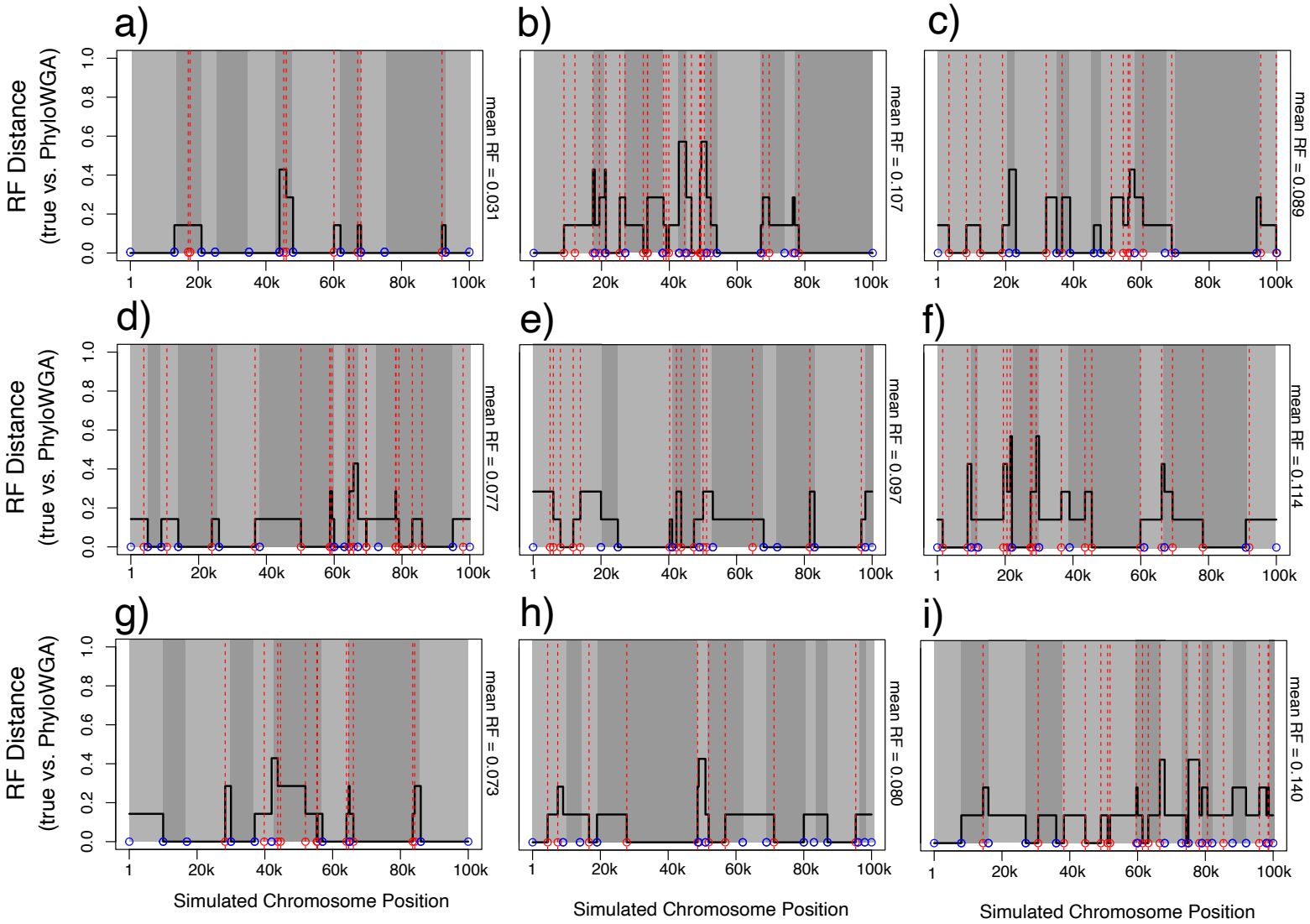

Figure S2. Results of the simulation analyses for demonstrating the accuracy of *PhyloWGA* on chromosome alignments simulated with a recombination rate of $r = 10^{-9}$ per site per generation. Each of nine replicate analyses are shown in each successive panels a-i. Red dashed lines indicate recombination events yielding differences in topology, while alternating light and dark gray blocks indicate concatenated windows from *Chromo.Crawl* with one kb windows and step sizes. Blue circles demarcate the boundaries of concatenated windows. Black lines measure the Robinson-Foulds (RF) distance between the true, simulated tree and the inferred tree from *Chromo.Phylome*.

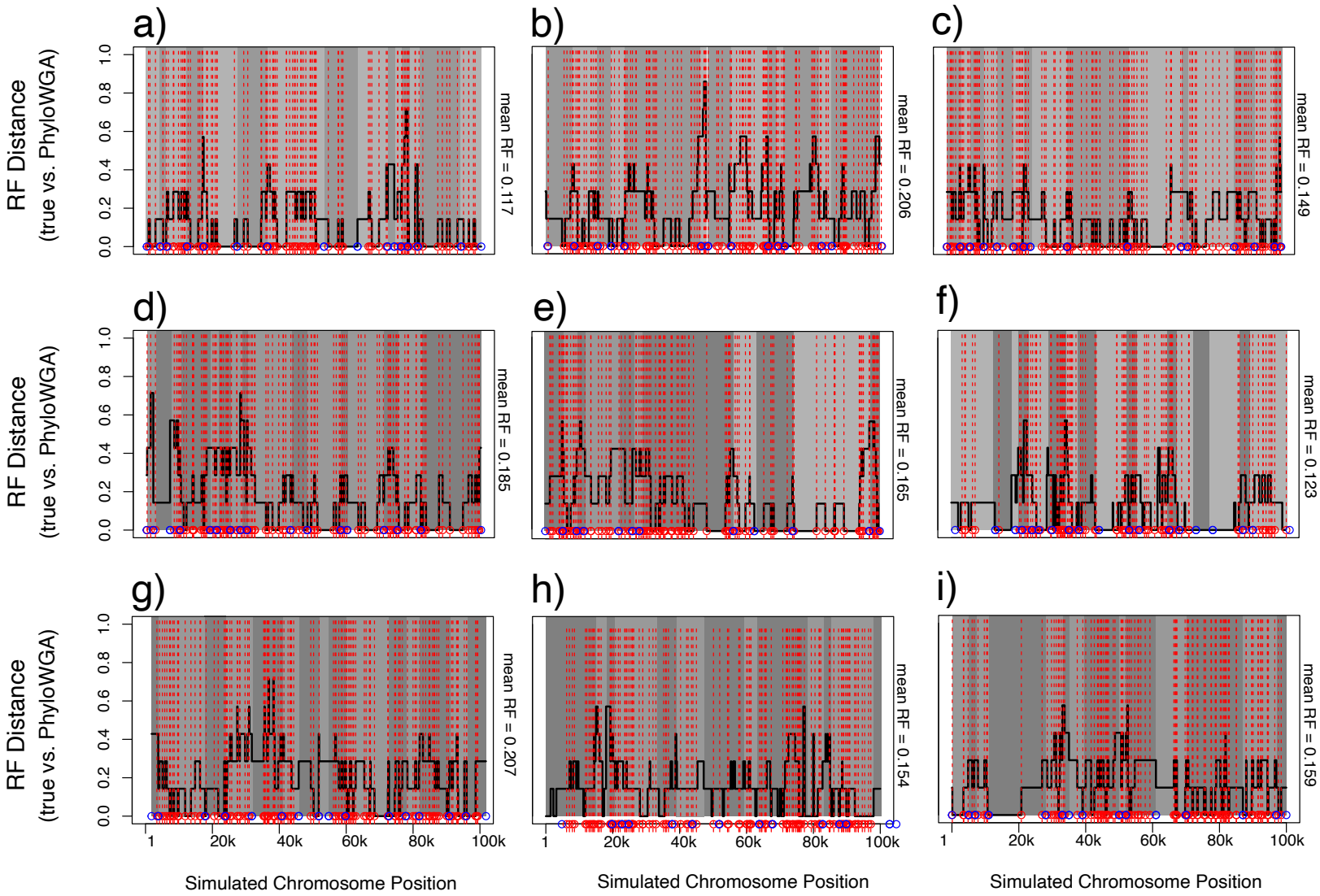

Figure S3. Results of the simulation analyses for demonstrating the accuracy of *PhyloWGA* on chromosome alignments simulated with a recombination rate of $r = 10^{-8}$ per site per generation. Each of nine replicate analyses are shown in each successive panels a-i. Red dashed lines indicate recombination events yielding differences in topology, while alternating light and dark gray blocks indicate concatenated windows from *Chromo.Crawl* with one kb windows and step sizes. Blue circles demarcate the boundaries of concatenated windows. Black lines measure the Robinson-Foulds (RF) distance between the true, simulated tree and the inferred tree from *Chromo.Phylome*.

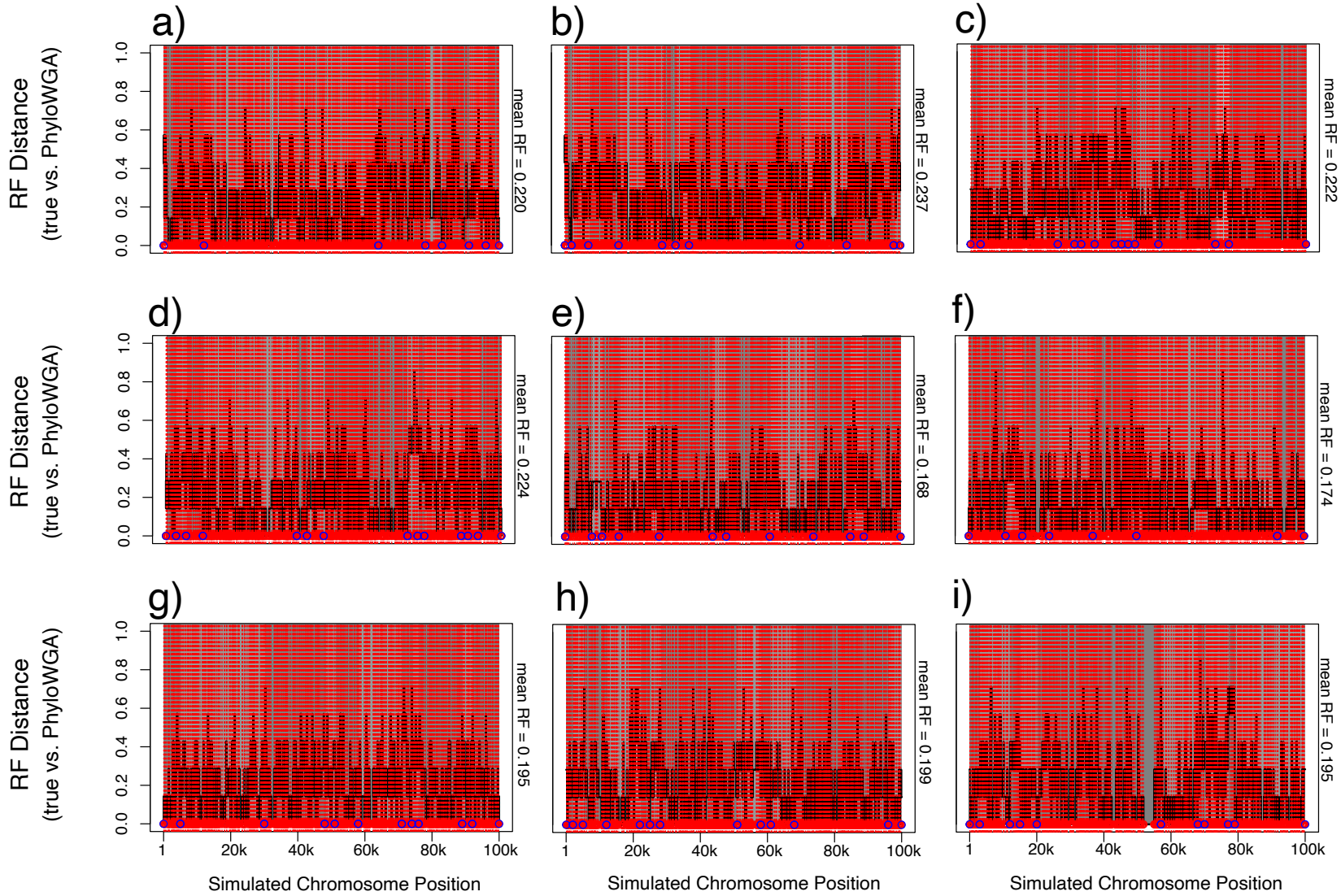

Figure S4. Results of the simulation analyses for demonstrating the accuracy of *PhyloWGA* on chromosome alignments simulated with a recombination rate of $r = 10^{-7}$ per site per generation. Each of nine replicate analyses are shown in each successive panels a-i. Red dashed lines indicate recombination events yielding differences in topology, while alternating light and dark gray blocks indicate concatenated windows from *Chromo.Crawl* with one kb windows and step sizes. Blue circles demarcate the boundaries of concatenated windows. Black lines measure the Robinson-Foulds (RF) distance between the true, simulated tree and the inferred tree from *Chromo.Phylome*.

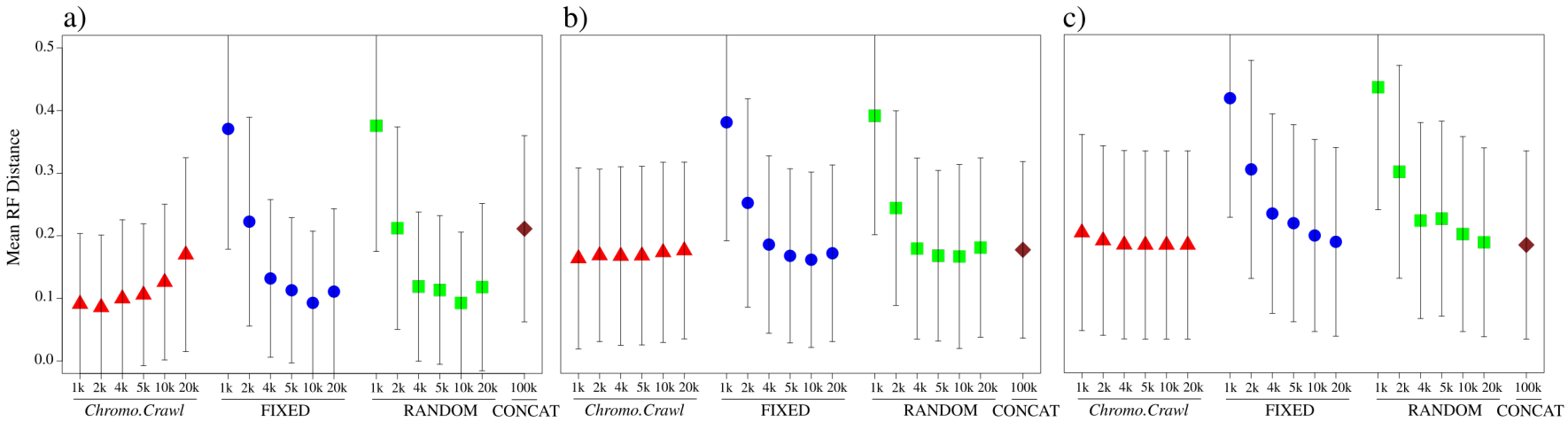

Figure S5. The adaptive window size for estimating gene trees provides *Chromo.Crawl* with improved chromosome-scale phylogenetic accuracy compared to typical approaches that do not approximate optimal window sizes. Phylogenetic accuracy of *Chromo.Crawl* (red triangles) compared with three alternative strategies: the FIXED approach using a fixed sliding window size to infer gene trees (blue circles), the RANDOM approach with trees inferred across randomly sampled genomic regions also of fixed size (green squares), and the CONCAT approach that assumes all windows share the same tree by concatenating the entire 100kb window (brown diamonds). Results shown for mean Robinson-Foulds (RF) distance (bars indicate standard deviation) across nine replicates for the (a) low ($r = 10^{-9}$), (b) medium ($r = 10^{-8}$) , and (c) high ($r = 10^{-7}$) recombination rate simulations using one, two, four, five, 10, and 20 kb window sizes, respectively (left to right in each panel).