Horizontal Gene Transfer Inference: Gene presence-absence outperforms gene trees

Swastik Mishra[1] **and Martin J. Lercher**[1] ✉

[1]Institute for Computer Science and Department of Biology, Heinrich Heine University, Düsseldorf, Germany

Horizontal gene transfer (HGT) is a fundamental driver of prokaryotic evolution, facilitating the acquisition of novel traits and adaptation to new environments. Despite its importance, methods for inferring HGT are rarely systematically compared, leaving a gap in our understanding of their relative strengths and limitations. Validating HGT inference methods is challenging due to the absence of a genomic fossil record that could confirm historical transfer events. Without an empirical gold standard, new inference methods are typically validated using simulated data; however, these simulations may not accurately capture biological complexity and often embed the same assumptions used in the inference methods themselves. Here, we leverage the tendency of HGT events to involve multiple neighboring genes to assess the accuracy of diverse HGT inference methods. We show that methods analyzing gene family presence/absence patterns across species trees consistently outperform approaches based on gene tree-species tree reconciliation. Our findings challenge the prevailing assumption that explicit phylogenetic reconciliation methods are superior to simpler implicit methods. By providing a comprehensive benchmark, we offer practical recommendations for selecting appropriate methods and indicate avenues for future methodological advancements.

Horizontal gene transfer, HGT inference, gene presence-absence profiles, reconciliation models, implicit phylogenetic methods, explicit phylogenetic methods

Correspondence: ***martin.lercher@hhu.de***

Introduction

Horizontal Gene Transfer (HGT) is recognized as a major driving force in the evolution of prokaryotes. (1) However, the inference of HGT events is challenging, since HGT is commonly followed by other evolutionary processes such as sequence amelioration and gene loss. Going back in time to validate any inferred events against a set of true HGT events is currently impossible. (2) In the best-case scenario, HGT events can be traced between closely related genomes in lab evolution experiments to validate HGT inferences. This limits any rigorous validation of HGT inference methods to only extremely recent HGT events. However, on longer timescales, transfers occur frequently even between distantly related organisms (3, 4), something that laboratory evolution experiments cannot replicate. Although one can test the robustness of the HGT inference methods, for example by data resampling, that does not validate the model choices made during the inference process.(5)

As a result of these challenges, HGT inference methods are typically validated using simulated data. Simulations are based on a set of assumptions about the evolutionary process, potentially biasing the simulations in comparison to real data – even if these assumptions are accepted principles or "natural interpretations" of the community (6). Examples of potentially biasing assumptions are: considering genes as evolutionary units, ignoring intragenic rearrangements; neglecting population-level processes such as drift; assuming that sites evolve independently from each other and from structural constraints; and ignoring extinct or unsampled species. Direct comparisons between HGT inference methods are complicated by the differences in assumptions made by each method, so each method may excel in specific simulation scenarios. Generally speaking, a particular HGT inference method will perform well in the simulated world designed for it. Methods are not tested in the real world, and they may thus fail to capture real-world complexities (7). The vast majority of methods, when published, have not even been tested on a common reference set of simulated data, nor have they been compared to one another. We believe that it is more important to develop some way of validating and comparing methods on real data than to refine existing methods or invent new ones without any explicit validation against biological reality.

In the following, we develop a comparative analysis of HGT inference methods using genomics data. Our goal is to provide an approximate ranking of HGT inference methods, delineating which methods infer results that are meaningful in a given biological context, as well as to have a nuanced view of how these methods compare to each other.

HGT inference methods can be broadly classified into two categories: *parametric* (or sequence composition) methods and *phylogenetic* (or tree-based) methods (8).

Parametric methods operate under the assumption that the *sequence composition* of the transferred gene differs between the donor and recipient organisms. Sequence features of a genome, such as GC content, codon usage, and oligonucleotide frequencies, are often analyzed. However, they are limited in their effectiveness when the donor and recipient are closely related and thus have similar sequence features, or when dealing with ancient HGT events where enough time has passed for the acquired nucleotide sequence to evolve to the sequence features of the host genome (sequence amelioration).

Phylogenetic methods, on the other hand, infer information across a set of sequences by using phylogenetic trees. These methods can be further divided into implicit and explicit approaches. Implicit phylogenetic methods include BLAST hit-based methods (e.g., DLIGHT (9)); however, the most

popular implicit methods are phyletic pattern-based (e.g., GLOOME (10), Count(11)). The latter use a maximum parsimony (MP) or maximum likelihood (ML) framework to infer *gain* and *loss* events that explain the observed gene distribution across genomes in a genome tree.

In contrast, explicit methods, including RANGER-DTL (12) and ALE (13, 14), compare gene trees with genome (or species) trees. They aim to infer the sets of Duplication, Transfer, and Loss events (DTL) that maximize the likelihood of the topology of a gene tree given that of the corresponding genome tree. One may then find one or more such *reconciliation* scenarios for a given gene tree / genome tree pair.

Although phylogenetic methods offer valuable insights, they are not without limitations. Unrecognized paralogy resulting from duplication followed by loss can be misinterpreted as HGT. Furthermore, tree reconstruction errors, particularly in gene trees, can lead to false HGT inferences (15).

Parametric methods were developed first and have been benchmarked elsewhere (16), with oligonucleotide signature-based methods found to outperform alternative approaches. Here, we thus focus primarily on the comparison of phylogenetic methods, using the oligonucleotide signature-based method Wn as a reference point. With increases in computing power, studies on HGT have increasingly used phylogenetic methods, not least because they are not compromised by sequence amelioration. They are also useful for not only detecting the presence of HGT in a gene's history, but also provide information on the phylogenetic branches on which genes were acquired and – for explicit methods – the likely donor of the sequence. Explicit methods included in our study are ALE, RANGER-DTL, and AnGST (17). For RANGER-DTL, in addition to the full maximum-likelihood version, we also include a heuristic 'fast' version that samples a smaller DTL parameter space. Implicit methods included in our study are GLOOME and Count, two phyletic pattern-based methods. GLOOME and Count were each run with both maximum parsimony (MP) and maximum likelihood (ML) settings. With GLOOME, each of those were also run with and without an input species tree; without an input tree, GLOOME identifies a tree topology most consistent with the phyletic patterns. Unless stated otherwise, we refer to GLOOME (ML or MP) as versions with an input species tree.

Our selection of HGT inference methods for the benchmark aims to provide a comprehensive overview of different state-of-the-art methods. We preferentially include methods that are widely used by the community and have user-friendly, publicly available implementations. There are many other methods that have not been included in this study. These include methods that have no publicly available implementations of the algorithm (e.g., DLIGHT (9)); methods that cannot handle modern large datasets with multiple genes of the same gene family in a taxon (especially subtree-prune-and-regraft based methods, e.g. RIATA-HGT (18), SPRIT (19), EEEP (20)); methods that focus on the presence of HGT in the evolution of a gene but not on the inference of exact transfer events (e.g. JPrime (21); and additional methods discussed in (22)).

Here, we benchmark the selected methods by measuring their reproduction of expected patterns in empirical data. Although the usage of explicit gene tree information may seem to be an advantage, our findings indicate that implicit phylogenetic methods infer more biologically meaningful HGT events than explicit phylogenetic methods or parametric methods, having a lower false positive rate.

Results

For our benchmark of HGT inference methods, we used the dataset of orthologous gene families in Gammaproteobacteria in the EggNOG database v6 (23). EggNOG terms these gene families non-supervised orthologous groups, or *NOGs*. We focussed on the 359 Gammaproteobacterial taxa in EggNOG that are featured in the tree of life provided by ASTRAL (24); we used the subtree with these taxa as the input genome tree for the phylogenetic methods. We then sampled a representative set of 1286 gene families (NOGs) that occur in at least 10 of these taxa. We obtained the corresponding gene trees from EggNOG, rooting them using Minimum Ancestor Deviation (25), which has been shown to be more accurate than other rooting methods (26). We used these gene families to infer HGT events using each of the tested methods.

Most of the tested methods have a tunable parameter that affects the number of HGT inferences for a given input. Here we define *transfer thresholds* based on these parameters. For the explicit phylogenetic methods, we set the transfer thresholds on the mean number of transfers inferred per reconciliation model. For GLOOME (ML), the threshold is on the minimal posterior probability of gene gain events. For Count (ML), we set transfer thresholds on the expected number of gene gain events. For Count (MP) the threshold is the gain penalty ratio. For the parametric method Wn, we fixed the window size to 8 nucleotides and used thresholds based on multiples of *median absolute deviation* (MAD, see Methods for details). These metrics are, in each case, a measure of the confidence of the inference method. However, they are not directly comparable across methods.

Implicit and explicit methods infer very different sets of HGT events. To assess similarities and differences in the sets of HGT events inferred by each method, we first clustered the methods based on the similarity between pairs of these sets. As a similarity measure, we used the Overlap Coefficient, which is the ratio of the size of the intersection to the size of the smaller set in the pair. Because for each method the set of infered HGT events can expand or contract massively by changing the transfer threshold, we measured similarities as the maximum possible Overlap Coefficient across all possible pairs of thresholds.

As expected, the explicit phylogenetic methods (ALE, RANGER, AnGST) cluster together, as do the implicit phylogenetic methods (GLOOME, Count) (Figure 1). The biggest differences in terms of inferred HGT events are seen between explicit and implicit phylogenetic methods. The paramet-

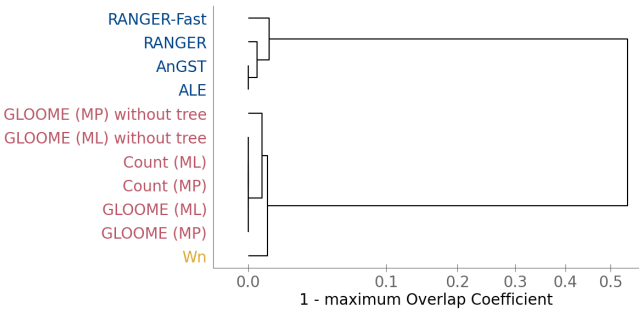

Fig. 1. Clustering analysis of the inferred sets of HGT events. Shown here is a UPGMA dendrogram of the tested methods based on their maximum Overlap Coefficient. Overlap Coefficient is the ratio of the size of the intersection to the size of the smaller set (of inferred HGT events) in the pair of methods. Maximum Overlap Coefficient is the maximum of this metric across all transfer thresholds of any pair of methods. This is a similarity measure between 0 and 1, and 1 minus this value is used here as a distance measure. Red: implicit phylogenetic methods. Blue: explicit phylogenetic methods. Yellow: parametric method Wn.

ric method Wn clusters with the implicit phylogenetic methods. Note that the maximum Overlap Coefficient between all the implicit methods is 1 (or very close to 1 when involving GLOOME (MP) without tree), i.e., for each pair of methods, there are thresholds such that one set of HGT inferences is a subset of the other. The same is true for ALE and AnGST.

Implicit phylogenetic methods infer a higher percentage of co-acquisitions as neighbors. We call two genes co-acquisitions if they were acquired by one genome on the same branch of the genome tree. If they were also acquired from the same donor genome, we call them co-transfers. The unit of HGT is a genomic fragment that may contain multiple genes, and thus many pairs of co-transferred genes may result from a single HGT event. Such gene pairs were neighbors in the donor genome and will – at least initially – also be neighbors in the recipient genome (27, 28). Later genomic rearrangements may break up this neighborship. Although two genes acquired on the same branch of the genome tree in independent HGT events can also be genomic neighbors, the likelihood of this happening by chance is low (27).

We have no strong *a priori* expectation about the relative probability that two co-acquired genes resulted from the same HGT event. However, if we see notable differences between methods in the fraction of inferred co-acquisitions that are genomic neighbors, then this indicates *a posteriori* that (i) some co-acquisitions indeed result from a single HGT event and that (ii) the method inferring a higher fraction of neighboring co-acquisitions has a lower false-positive rate.

For each method, the number of inferred co-acquisitions varied based on the transfer threshold used for each inference method. As we decrease the transfer threshold, the number of co-acquisitions inferred by ALE, Wn, Count, and GLOOME (ML) varies from about 100 to 5×10^5. In contrast, GLOOME (ML, without tree), and RANGER infer much higher numbers of co-acquisitions in general, varying from about 2×10^4 to 10^6 (Supplementary Figure S1c). For a fair comparison, the percentage of neighbors should be compared at comparable levels of stringency, i.e., at the same number of inferred co-acquisitions. For the main text, we define two genes as neighbors if they are separated by at most one other gene. The Supplementary Table S1 summarizes the results for other values of the maximum number of intervening genes.

Method	Mean	σ	Max	Threshold
Count (MP)	1.92825	1.61469	4.97238	7.00000
GLOOME (ML)	1.05923	0.47908	1.84000	0.95730
Count (ML)	0.97052	0.59988	2.60000	0.99999
Wn	0.59124	0.56239	1.61290	13.00000
ALE	0.50029	0.14463	0.80000	0.86000
GLOOME (MP)	0.47607	0.00000	0.47607	1.00000
GLOOME (ML)$^\alpha$	0.42425	0.06465	0.51257	0.74680
GLOOME (MP)$^\alpha$	0.41949	0.00000	0.41949	1.00000
AnGST	0.33321	0.00000	0.33321	1.00000
RANGER-Fast	0.33004	0.00132	0.33204	1.00000
RANGER	0.32130	0.03871	0.41729	1.00000

Table 1. Count (MP) infers a higher corrected percentage of neighboring co-acquisitions. Shown here are the mean, standard deviation, and maximum of the corrected percentage of neighboring co-acquisitions, as well as the transfer threshold at which the maximum occurs. This table shows only $t = 1$, where t is the maximum number of intervening genes between two co-acquired genes to be considered neighbors. Supplementary Table S1 shows the same for higher values of t. Methods using presence-absence outperform the others for all tested values of t, with Count (MP) leading among them. Methods such as ALE or Wn have lower variation across their transfer thresholds and low mean percentages of neighboring co-acquisitions. α : GLOOME without species tree

We defined a corrected percentage of neighboring co-acquisitions as the inferred percentage of neighboring co-acquisitions minus random expectations (i.e., the percentage expected if the co-acquisitions were randomly distributed across the genome; see Methods).

At any given number of co-acquisitions – i.e., any given level of stringency – the implicit phylogenetic methods Count (ML or MP) and GLOOME (ML or MP) almost always infer a higher corrected percentage of neighboring co-acquisitions than the explicit phylogenetic methods or Wn (Figure 2, Table 1). For example, at 10^3 co-acquisitions, Count (MP) infers more than 8 times the number of corrected neighboring co-acquisitions inferred by ALE. Across transfer thresholds, Count (MP) outperformed all other methods, with GLOOME (ML) taking second place except at the most stringent thresholds. GLOOME and Count rely solely on the presence and absence of genes across the genome tree.

It appears likely that gene trees are more accurate when they are based on more sequence information, that is, for longer gene sequences and larger gene trees. In this case, the methods that utilize gene tree information should perform better on longer genes or larger gene trees. However, when we limit the analysis to longer genes (Figure S2) or larger gene trees (Figure S3), the explicit phylogenetic methods still infer lower corrected percentages of neighboring co-acquisitions than Count (MP), GLOOME (ML) or Count (ML) (in the latter case with the exception of medium length genes when ALE outperforms Count (ML) (Figure S2b)). In sum, the use of gene trees appears to lead to a higher fraction of false positive HGT inferences.

We note here that given the small percentages of neighboring co-acquisitions, one must be cautious in interpreting the re-

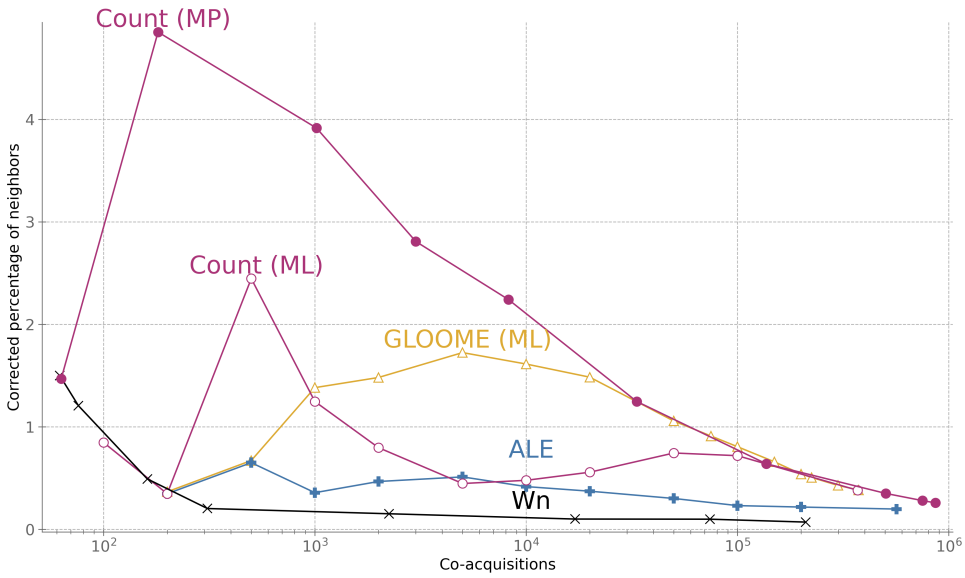

Fig. 2. Count (MP) outperforms other methods at inferring neighboring co-acquisitions. The y-axis shows the percentage of neighboring co-acquisitions inferred, minus that expected by random chance. More stringent transfer thresholds of individual methods are expected to lead to fewer false positive predictions; to make results comparable across methods, the x-axis indicates the number of co-acquisitions. Additional methods are shown in Supplementary Figure S1c.

sults. For example, in Figure S1b, the observed percentage of neighboring co-acquisitions for several methods (e.g., ALE, Count (ML), Wn) increases from 0.5% to 1% when we go from about 200 co-acquisitions to 100 co-acquisitions. This means that the same, single neighboring co-acquisition is inferred by the respective method, but the total number of co-acquisitions inferred is halved. Such effect of small numbers greatly affects the apparent performance of the methods. This type of effect also explains why the parametric method Wn, which generally performs worse than all other methods, appears to improve in performance at very high transfer thresholds (Figure 2).

The explicit phylogenetic methods in our analysis — ALE, RANGER, and AnGST – also provide information on the donors of the acquired genes. Genes acquired on the same branch of the genome tree could originate from the same donor or from different donors. We have no strong *a priori* expectation about the relative probability of these two types of events. However, it appears unlikely that a method would infer the same donor just by chance. Thus, if we see notable differences between methods in the fraction of inferred co-acquisitions that are also inferred to be co-transfers, then this indicates *a posteriori* that (i) some co-acquisitions are indeed co-transfers and that (ii) the method inferring a higher fraction of co-transfers has a lower false-positive rate. Thus, just like for neighboring co-acquisitions, we also expect that more reliable methods should infer more co-transfers among co-acquisitions; further, they should also infer more neighboring co-transfers among co-acquisitions.

Where results are available for both explicit phylogenetic methods, RANGER almost always infers higher percentages of co-transfers among co-acquired genes than ALE (Figure 3a). At increasing stringencies for ALE – where no RANGER results are available – ALE infers higher percentages of co-transferred genes.

In terms of the inferred percentage of co-transferred neighbors among co-acquisitions, RANGER and ALE perform similarly (Figure 3b). The results for ALE at higher stringency – where we have no data from RANGER – can again be attributed to the effects of small numbers. Given that the performance of the two methods in this test is similar, one may choose ALE over RANGER simply because it is faster to run. In our analyses, ALE finished running on the dataset using 50 threads in $\approx$ 23 minutes, whereas RANGER took $\approx$ 8 hours. AnGST and RANGER-Fast, while faster, perform much worse: they consistently infer very similar percentages of co-transferred genes and co-transferred neighbors, which are substantially lower than those inferred by ALE and RANGER.

Implicit methods provide the strongest evidence for increased transferability of response genes. HGT is generally seen as an important mechanism by which bacteria respond to environmental changes. Thus, we expect that genes involved in environmental responses are transferred more frequently than many other genes. We identified such *response genes* using Gene Ontology (GO) annotations (29). The number of HGT events inferred for a given gene family also depends on the gene family's size (number of orthologous genes in the *NOG*; see Suppl. Table S2). To control for this effect, we calculated the relative excess in the number of transfers inferred for a given gene family over the expected number for a gene family of its size; we term this number *transferrability* (see Methods). We then calculated the increase in transferability for response genes relative to other genes. Similar to the analyses of co-acquisitions, at comparable numbers of inferred HGTs, one can compare this increase in transferability between methods.

For a given method, the increase in inferred transferability of response genes is statistically significant only at some transfer thresholds (Figure S6; Supplementary Table S3 shows the Mann-Whitney U statistics, sample sizes, means, and p-values of the tests). When restricting our analysis to transfer thresholds where the increase in transferability is statistically significant (Figure 4), GLOOME (ML) and Count (ML) in-

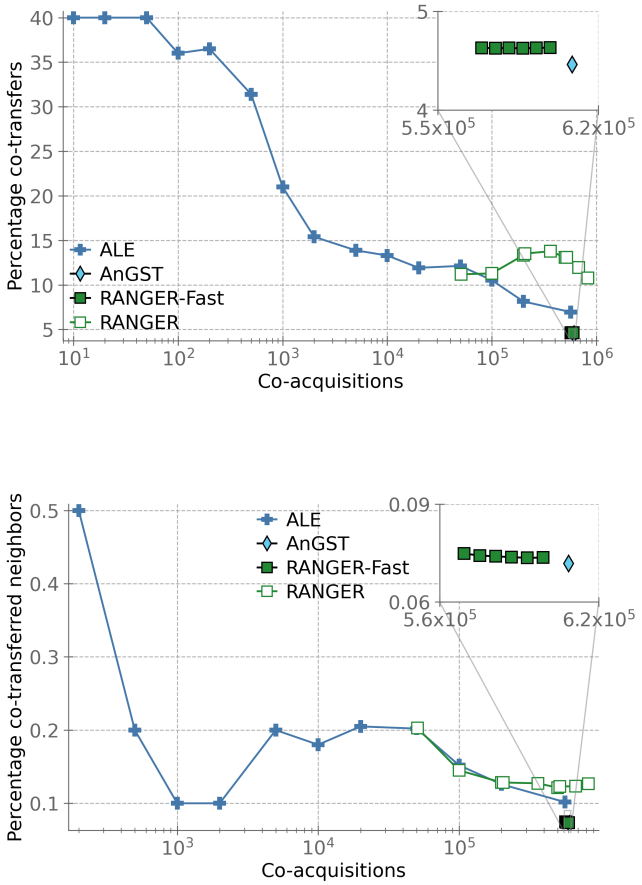

Fig. 3. RANGER performs better or similar to ALE at inferring co-transferred genes and co-transferred neighbors. At lower stringency, RANGER infers higher (a) percentages of co-transfers and (b) percentages of co-transferred neighbors than ALE. Although at higher stringency, ALE generally makes better inferences, it is either similar to RANGER or the performance can not be compared as no data from RANGER is available. AnGST and RANGER-Fast generally provide much worse inferences than ALE and RANGER. Note that only explicit phylogenetic methods are shown here, since other methods do not provide information about the source of transfer.

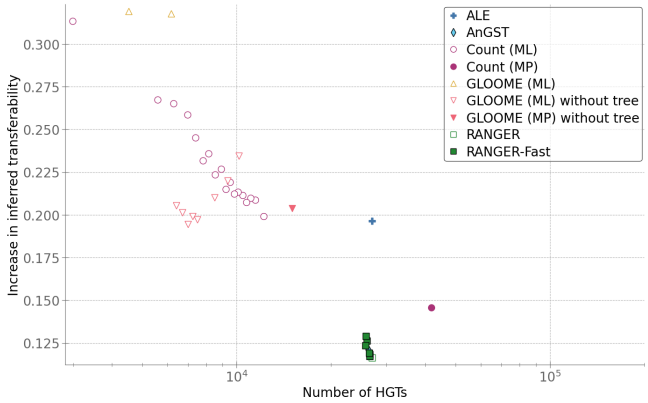

Fig. 4. GLOOME (ML) and Count (ML) outperform other methods at inferring response genes as highly transferred. Shown here are the increases in the transferability of response genes compared to non-response genes for methods at transfer thresholds where this increase is statistically significant (two-tailed $p < 0.05$, Mann-Whitney U-tests, see Supplementary Table S3). For all other methods as well as other transfer thresholds, see Supplementary Figure S8.

fer higher increases in transferability for response genes than

Method	Mean increase in transferability	Max p-value across transfer thresholds
GLOOME (ML)	0.318	0.0429
Count (ML)	0.231	0.047
GLOOME (ML)$^{\alpha}$	0.208	0.0458
GLOOME (MP)$^{\alpha}$	0.204	0.0154
ALE	0.196	0.00522
Count (MP)	0.146	0.0497
RANGER-Fast	0.122	0.015
AnGST	0.121	0.00997
RANGER	0.116	0.0111

Table 2. GLOOME (ML) and Count (ML) infer the highest mean increases in transferability of response genes. Shown here is the mean increase in inferred transferability for response genes, compared to non-response genes, for methods at transfer thresholds where this increase is statistically significant (two-tailed $p < 0.05$, Mann-Whitney U-tests, see Supplementary Table S3). α : GLOOME without species tree

other methods (Table 2). Surprisingly, Count (MP) performs quite poorly in this analysis. However, the definition of what is considered a response gene is not based on quantitative data, but on the GO annotations and the *ad hoc* keywords used here. We therefore place less weight on this analysis compared to the co-acquisition analyses discussed above.

Discussion

The present study is the first systematic comparison of phylogenetic methods for inferring HGT. For this, we introduce a data-driven benchmark, which we argue offers a more reliable standard than simulations. Our analysis reveals several important insights that can guide the choice of method for HGT inference in a given biological study.

Implicit phylogenetic methods, which rely solely on phylogenetic profiles, better capture the spatial clustering of horizontally transferred genes in recipient genomes. We suspect that the worse performance of explicit phylogenetic methods may lie in their reliance on gene trees, which may often not reflect the true evolutionary history of a gene family. These methods infer HGT events based on discrepancies between the topology of a gene tree and that of the corresponding species tree; therefore, gene tree reconstruction errors may lead to more false positive acquisitions and co-acquisitions. This is in line with previous studies that have shown that gene tree reconstruction errors can lead to false positives in HGT inference (15). The conjecture that noise in gene tree reconstruction is responsible for a lower reliability of explicit phylogenetic methods is further supported by the slight improvements in the performance of ALE and RANGER when the analyses are limited to longer genes or larger gene trees.

It is surprising to find Count (MP) performing on par or often better than Count (ML) and GLOOME (ML), given that maximum-likelihood is often considered a superior approach, e.g., in phylogeny reconstruction (30). The difference between the two versions of Count is larger when the transfer threshold is higher. The strong increase in the percentage of neighboring co-acquisitions indicates that the strategy of increasing Count (MP)'s stringency by increasing the gain/loss penalty ratio works well, leading to fewer, but more accurate

HGTs. In contrast, increasing the stringency of maximum-likelihood inferences by selecting events with increasingly higher inferred posterior probabilities (GLOOME) or expectation values (Count) appears to be less effective. Future work should explore alternative ways to vary the stringency of maximum-likelihood inferences.

Among tree reconciliation methods, RANGER and ALE generally performed better or similar to the others at inferring co-transfers and co-transferred neighbors. Given that ALE is aware of extinct taxa (13), it is surprising that it does not perform better than RANGER where results are available for both.

At first glance, the sequence composition-based method, Wn, can in theory benefit from direct access to sequence information. However, our results suggest that it requires very high stringency to perform in line with other methods; even there, Wn's apparent improvement in performance may be due to the small numbers of inferred neighbors.

Genes involved in environmental response are often subject to frequent HGT due to their role in adaptation to new niches. Our analysis revealed that GLOOME (ML) and Count (ML) excel at inferring these genes as highly transferred compared to others.

Keeping these results in mind, we recommend the use of the implicit phylogenetic methods Count (MP or ML) or GLOOME (ML) to identify horizontal gene acquisitions. In particular, Count (MP) outperforms all other methods in identifying neighboring co-acquisitions, especially at higher stringencies (gain/loss penalty ratios 4 to 7; Fig. 2). If inferring the donor of HGT events is a requirement, explicit phylogenetic methods must be used. In this case, both ALE and RANGER are adequate choices, although ALE is computationally more efficient than RANGER. ALE can be made more stringent than RANGER by choosing a high transfer threshold (> 0.86 transfers per reconciliation model). At this level of stringency, its inferences are the most reliable among all tested scenarios for explicit methods, as indicated by the highest percentages of neighboring co-acquisitions or co-transfers (Fig. S1c). On that note, for any method used, we generally recommend using a minimum transfer threshold (see Table 1) to increase stringency and focus on a subset of reliable HGT inferences.

Although Gammaproteobacteria are a well-studied and diverse bacterial clade, these findings might vary for other datasets. Specifically, results could differ when analyzing only recent gene transfers or examining shallower phylogenetic trees, such as those limited to *Escherichia coli* strains. Because many users will not spend time optimizing most of the parameters offered by the various programs, we ran each program only with their default settings. We note, however, that these parameters can dramatically change their performance.

Our conclusions are based predominantly on the analysis of neighboring co-acquisitions and co-transfers. Although it is challenging to establish an *a priori* expectation for the likelihood that two co-acquired genes originate from the same HGT event, the notable differences between methods in the fraction of inferred co-acquisitions that are genomic neighbors suggest that some co-acquisitions do indeed result from a single HGT event. This provides strong evidence that methods inferring a higher fraction of neighboring co-acquisitions have a lower false positive rate. If no such signal existed, we would expect similar percentages of neighboring co-acquisitions, as well as similar proportions of co-transferred genes and co-transferred neighbors, across methods. Since this is not the case, we can conclude that certain methods are more reliable than others. Consequently, this benchmark should aid in the development of improved HGT inference methods in the future.

Methods

Genomic data selection and cleaning. We used the Gammaproteobacterial dataset (NCBI taxonomic ID: 1236) from the EggNOG database v6 (23) for orthologous genes (non-supervised orthologous groups, or *NOGs*). The corresponding nucleotide sequence information (for Wn execution) was downloaded from the NCBI GenBank database (31) using the NCBI Datasets CLI program. The reference species tree downloaded from ASTRAL WoL (24). Specifically, the `astral - branch length - cons - astral.cons.nwk` file was used. In brief, this means that the topology of the tree was inferred using ASTRAL and the branch lengths were re-estimated with maximum likelihood, using the 100 most conserved sites per gene. We retain only the taxa belonging to the maximum overlap of taxa between the set of NOGs, such that their sequence data could be retrieved and that these taxa could be reliably mapped onto the species tree. Furthermore, to ensure that gene families were large enough to infer meaningful HGT events, NOGs containing less than 10 genes or 30 taxa were removed.

NOGs were then selected around the average sized NOGs such that they together represented all the 359 taxa in the dataset, resulting in a set of 1286 NOGs. The gene trees were taken from the EggNOG database and rooted using Minimum Ancestor Deviation (25).

Inferring HGT. All the programs were executed separately, using default parameters. For ALE, the *ALEml_undated* program was used, which doesn't require a dated species tree (32). RANGER-DTL v2.0 was executed with and without the 'fast' option. AnGST was executed with parameters as suggested in the manual. For Count, the 'Asymmetric Wagner Parsimony' method was used as the maximum parsimony method. Unlike GLOOME where we also ran without a species tree for both ML and MP settings, Count could be run only with a species tree.

The parametric method Wn was implemented based on the original publication (33), and as suggested, we used an oligonucleotide size of 8 nucleotides. Wn calculates a *typicality score* of how similar the gene sequence is to the genome sequence based on the oligonucleotide frequencies in both. These frequencies were efficiently calculated using KMC (34). In the original paper for Wn, the authors used an automatic threshold detection method. This method rank or-

ders the genes based on their typicality scores, smoothens this curve (requiring a rolling window average), takes a derivative of this curve, and then considers transferred genes as those that are above the point where the derivative becomes *approximately* constant. Although this method is termed *automatic*, it still depends on both the rolling window size used for smoothing and the user's notion of when exactly does the derivative become approximately constant (i.e. the double derivative is approximately zero). Instead of making the choice of when the double derivative is approximately zero, we used multiples of the *median absolute deviation* (MAD) of the double derivative around zero, which also helps to set the transfer thresholds. These multiples represent a *stringency* parameter that goes from 3 times the MAD to 12 times the MAD. Note that it is generally recommended to use at least 2.5 times the MAD, and so we are already using a relatively high stringency even when Wn is run at the lowest stringency. The rolling window size was automatically determined by iteratively finding the longest Savitzky-Golay filter such that the root mean square error of the smoothed curve with respect to the original curve stops decreasing.

Co-acquisition and co-transfer analysis. The expected percentage of neighboring co-acquisitions is calculated as follows: if t is the max. number of intervening genes between the two genes in the pair (in our case t=1), and n is the number of chromosomes among the inferred co-acquisitions for a method, the expected fraction of neighboring co-acquisitions is

$$\frac{2\sum_{i=1}^{n}(t+1)(c_i/g_i)}{\sum_{i=1}^{n}c_i}$$

where c_i and g_i are the number of co-acquisitions and the number of genes in chromosome i, respectively. This is because for each horizontally transferred gene in a chromosome, the probability of another gene being transferred as a neighbor is $2(t+1)/g_i$ (i.e. approximating for boundary genes, there are $t+1$ neighboring positions on either side of a gene out of g_i positions in the chromosome). The expected number of neighboring co-acquisitions is then this probability multiplied by the number of co-acquisitions for that chromosome. The expected *fraction* of neighboring co-acquisitions is the sum of the expected number of neighboring co-acquisitions (per chromosome) divided by the sum of co-acquisitions for all chromosomes.

Due to the dependence on the exact chromosomes present in inferences for any given method, the expected percentage of neighboring co-acquisitions can vary between methods and across the number of inferred co-acquisitions for the same method (Figure S1a).

In any gene family, genes that could not be retrieved from GenBank, or pair of genes co-acquired by the same taxon but not in the same chromosome/contig in the GenBank data, were excluded from this analysis.

Modeling HGT in terms of gene function or connectivity. Gene Ontology (GO) annotations (29, 35) were used to identify gene function. There are 719 NOGs in our dataset for which GO annotations were available in EggNOG. To select response genes, the following keywords were searched for in the annotations: 'response', 'defense', and 'detoxification'. This resulted in 48 NOGs being classified as response gene families.

The modeling approach corrects for the gene family size being a confounding factor. First, we fit an ordinary least squares (OLS) regression model to express the inferred transfers for a given gene family as a function of the number of genes in the gene family. We then calculated the residuals of this fit, i.e., the distance of the number of transfers to the expected number given by the regression model. For each family, we then defined the increase in transferability, corrected for the gene family size, as the mean of the inferred transferability of response genes minus the mean of non-response genes. Two-tailed p-values for whether this increase in transferability is more extreme than expected by random chance were calculated using the Mann-Whitney U test.

Data availability

All of the code used in this study is available at the following Gitlab repository: https://gitlab.cs.uni-duesseldorf.de/general/ccb/hgt_inference_comparative_study The data used in this study can be retrieved from Zenodo: https://doi.org/10.5281/zenodo.14781414.

Competing interests

The authors have declared no competing interest.

Funding

This work was supported by a grant from Volkswagenstiftung in the "Life?" initiative (to M.J.L.)

Bibliography

1. Brian J. Arnold, I.-Ting Huang, and William P. Hanage. Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology*, pages 1–13, November 2021. ISSN 1740-1534. doi: 10.1038/s41579-021-00650-4.
2. J. Richard Gott. *Time Travel in Einstein's Universe: The Physical Possibilities of Travel Through Time*. HMH, August 2015. ISBN 978-0-547-52657-7.
3. Otto X. Cordero and Paulien Hogeweg. The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proceedings of the National Academy of Sciences*, 106(51): 21748–21753, December 2009. doi: 10.1073/pnas.0907584106.
4. Michael Sheinman, Ksenia Arkhipova, Peter F Arndt, Bas E Dutilh, Rutger Hermsen, and Florian Massip. Identical sequences found in distant genomes reveal frequent horizontal transfer across the bacterial domain. *eLife*, 10:e62719, June 2021. ISSN 2050-084X. doi: 10.7554/eLife.62719.
5. Joseph Felsenstein. *Inferring Phylogenies*. Sinauer, October 2003. ISBN 978-0-87893-177-4.
6. Paul Feyerabend. *Against Method*. Verso, 1993. ISBN 978-0-86091-646-8.
7. Nils Kapust, Shijulal Nelson-Sathi, Barbara Schönfeld, Einat Hazkani-Covo, David Bryant, Peter J. Lockhart, Mayo Röttger, Joana C. Xavier, and William F. Martin. Failure to Recover Major Events of Gene Flux in Real Biological Data Due to Method Misapplication. *Genome Biology and Evolution*, 10(5):1198–1209, May 2018. doi: 10.1093/gbe/evy080.
8. Matt Ravenhall, Nives Škunca, Florent Lassalle, and Christophe Dessimoz. Inferring Horizontal Gene Transfer. *PLoS Computational Biology*, 11(5), May 2015. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1004095.
9. Christophe Dessimoz, Daniel Margadant, and Gaston H. Gonnet. DLIGHT – Lateral Gene Transfer Detection Using Pairwise Evolutionary Distances in a Statistical Framework. In Martin Vingron and Limsoon Wong, editors, *Research in Computational Molecular Biology*, pages 315–330, Berlin, Heidelberg, 2008. Springer. ISBN 978-3-540-78839-3. doi: 10.1007/978-3-540-78839-3_27.
10. Ofir Cohen, Haim Ashkenazy, Frida Belinky, Dorothée Huchon, and Tal Pupko. GLOOME: Gain loss mapping engine. *Bioinformatics*, 26(22):2914–2915, November 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq549.

11. Miklós Csűös. Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*, 26(15):1910–1912, August 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq315.
12. Mukul S. Bansal, Manolis Kellis, Misagh Kordi, and Soumya Kundu. RANGER-DTL 2.0: Rigorous reconstruction of gene-family evolution by duplication, transfer and loss. *Bioinformatics*, 34(18):3214–3216, September 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty314.
13. Gergely J. Szöllosi, Bastien Boussau, Sophie S. Abby, Eric Tannier, and Vincent Daubin. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proceedings of the National Academy of Sciences of the United States of America*, 109(43):17513–17518, October 2012. ISSN 1091-6490. doi: 10.1073/pnas.1202997109.
14. Gergely J. Szöllősi, Eric Tannier, Nicolas Lartillot, and Vincent Daubin. Lateral Gene Transfer from the Dead. *Systematic Biology*, 62(3):386–397, May 2013. ISSN 1063-5157. doi: 10.1093/sysbio/syt003.
15. Cuong Than, Derek Ruths, Hideki Innan, and Luay Nakhleh. Confounding factors in HGT detection: Statistical error, coalescent effects, and multiple solutions. *Journal of Computational Biology*, 14:517–535, 2007.
16. Jennifer Becq, Cécile Churlaud, and Patrick Deschavanne. A Benchmark of Parametric Methods for Horizontal Transfers Detection. *PLoS ONE*, 5(4):e9989, April 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0009989.
17. Lawrence A. David and Eric J. Alm. Rapid evolutionary innovation during an Archaean genetic expansion. *Nature*, 469(7328):93–96, January 2011. ISSN 1476-4687. doi: 10.1038/nature09649.
18. Luay Nakhleh, Derek Ruths, and Li-San Wang. RIATA-HGT: A Fast and Accurate Heuristic for Reconstructing Horizontal Gene Transfer. In Lusheng Wang, editor, *Computing and Combinatorics*, Lecture Notes in Computer Science, pages 84–93, Berlin, Heidelberg, 2005. Springer. ISBN 978-3-540-31806-4. doi: 10.1007/11533719_11.
19. Tobias Hill, Karl JV Nordström, Mikael Thollesson, Tommy M Säfström, Andreas KE Vernersson, Robert Fredriksson, and Helgi B Schiöth. SPRIT: Identifying horizontal gene transfer in rooted phylogenetic trees. *BMC Evolutionary Biology*, 10:42, February 2010. ISSN 1471-2148. doi: 10.1186/1471-2148-10-42.
20. Robert G. Beiko and Nicholas Hamilton. Phylogenetic identification of lateral genetic transfer events. *BMC Evolutionary Biology*, 6(1):1–17, December 2006. ISSN 1471-2148. doi: 10.1186/1471-2148-6-15.
21. Joel Sjöstrand, Ali Tofigh, Vincent Daubin, Lars Arvestad, Bengt Sennblad, and Jens Lagergren. A Bayesian Method for Analyzing Lateral Gene Transfer. *Systematic Biology*, 63(3):409–420, May 2014. ISSN 1063-5157. doi: 10.1093/sysbio/syu007.
22. Maria Poptsova. Testing Phylogenetic Methods to Identify Horizontal Gene Transfer. In John M. Walker, Maria Boekels Gogarten, Johann Peter Gogarten, and Lorraine C. Olendzenski, editors, *Horizontal Gene Transfer*, volume 532, pages 227–240. Humana Press, Totowa, NJ, 2009. ISBN 978-1-60327-852-2 978-1-60327-853-9. doi: 10.1007/978-1-60327-853-9_13.
23. Jaime Huerta-Cepas, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K Forslund, Helen Cook, Daniel R Mende, Ivica Letunic, Thomas Rattei, Lars J Jensen, Christian von Mering, and Peer Bork. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(Database issue):D309–D314, January 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1085.
24. Qiyun Zhu, Uyen Mai, Wayne Pfeiffer, Stefan Janssen, Francesco Asnicar, Jon G. Sanders, Pedro Belda-Ferre, Gabriel A. Al-Ghalith, Evguenia Kopylova, Daniel McDonald, Tomasz Kosciolek, John B. Yin, Shi Huang, Nimaichand Salam, Jian-Yu Jiao, Zijun Wu, Zhenjiang Z. Xu, Kalen Cantrell, Yimeng Yang, Erfan Sayyari, Maryam Rabiee, James T. Morton, Sheila Podell, Dan Knights, Wen-Jun Li, Curtis Huttenhower, Nicola Segata, Larry Smarr, Siavash Mirarab, and Rob Knight. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nature Communications*, 10(1):5477, December 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-13443-4.
25. Fernando Domingues Kümmel Tria, Giddy Landan, and Tal Dagan. Phylogenetic rooting using minimal ancestor deviation. *Nature Ecology & Evolution*, 1(1):1–7, June 2017. ISSN 2397-334X. doi: 10.1038/s41559-017-0193.
26. Taylor Wade, L. Thiberio Rangel, Soumya Kundu, Gregory P. Fournier, and Mukul S. Bansal. Assessing the accuracy of phylogenetic rooting methods on prokaryotic gene families. *PLOS ONE*, 15(5):e0232950, May 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0232950.
27. Alexander Dilthey and Martin J. Lercher. Horizontally transferred genes cluster spatially and metabolically. *Biology Direct*, 10(1):72, December 2015. ISSN 1745-6150. doi: 10.1186/s13062-015-0102-5.
28. Tin Yau Pang and Martin J Lercher. Supra-operonic clusters of functionally related genes (SOCs) are a source of horizontal gene co-transfers. *Scientific reports*, 7:40294, January 2017. ISSN 2045-2322. doi: 10.1038/srep40294.
29. The Gene Ontology Consortium, Suzi A Aleksander, James Balhoff, Seth Carbon, J Michael Cherry, Harold J Drabkin, Dustin Ebert, Marc Feuermann, Pascale Gaudet, Nomi L Harris, David P Hill, Raymond Lee, Huaiyu Mi, Sierra Moxon, Christopher J Mungall, Anushya Muruganugan, Tremayne Mushayahama, Paul W Sternberg, Paul D Thomas, Kimberly Van Auken, Jolene Ramsey, Deborah A Siegele, Rex L Chisholm, Petra Fey, Maria Cristina Aspromonte, Maria Victoria Nugnes, Federica Quaglia, Silvio Tosatto, Michelle Giglio, Suvarna Nadendla, Giulia Antonazzo, Helen Attrill, Gil dos Santos, Steven Marygold, Victor Strelets, Christopher J Tabone, Jim Thurmond, Pinglei Zhou, Saadullah H Ahmed, Praoparn Asanitthong, Diana Luna Buitrago, Meltem N Erdol, Matthew C Gage, Mohamed Ali Kadhum, Kan Yan Chloe Li, Miao Long, Aleksandra Michalak, Angeline Pesala, Armalya Pritazahra, Shirin C C Saverimuttu, Renzhi Su, Kate E Thurlow, Ruth C Lovering, Colin Logie, Snezhana Oliferenko, Judith Blake, Karen Christie, Lori Corbani, Mary E Dolan, Harold J Drabkin, David P Hill, Li Ni, Dmitry Sitnikov, Cynthia Smith, Alayne Cuzick, James Seager, Laurel Cooper, Justin Elser, Pankaj Jaiswal, Parul Gupta, Pankaj Jaiswal, Sushma Naithani, Manuel Lera-Ramirez, Kim Rutherford, Valerie Wood, Jeffrey L De Pons, Melinda R Dwinell, G Thomas Hayman, Mary L Kaldunski, Anne E Kwitek, Stanley J F Laulederkind, Marek A Tutaj, Mahima Vedi, Shur-Jen Wang, Peter D'Eustachio, Lucila Aimo, Kristian Axelsen, Alan Bridge, Nevila Hyka-Nouspikel, Anne Morgat, Suzi A Aleksander, J Michael Cherry, Stacia R Engel, Kalpana Karra, Stuart R Miyasato, Robert S Nash, Marek S Skrzypek, Shuai Weng, Edith D Wong, Erika Bakker, Tanya Z Berardini, Leonore Reiser, Andrea Auchincloss, Kristian Axelsen, Ghislaine Argoud-Puy, Marie-Claude Blatter, Emmanuel Boutet, Lionel Breuza, Alan Bridge, Cristina Casals-Casas, Elisabeth Coudert, Anne Estreicher, Maria Livia Famiglietti, Marc Feuermann, Arnaud Gos, Nadine Gruaz-Gumowski, Chantal Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Philippe Le Mercier, Damien Lieberherr, Patrick Masson, Anne Morgat, Ivo Pedruzzi, Lucille Pourcel, Sylvain Poux, Catherine Rivoire, Shyamala Sundaram, Alex Bateman, Emily Bowler-Barnett, Hema Bye-A-Jee, Paul Denny, Alexandr Ignatchenko, Rizwan Ishtiaq, Antonia Lock, Yvonne Lussi, Michele Magrane, Maria J Martin, Sandra Orchard, Pedro Raposo, Elena Speretta, Nidhi Tyagi, Kate Warner, Rossana Zaru, Alexander D Diehl, Raymond Lee, Juancarlos Chan, Stavros Diamantakis, Daniela Raciti, Magdalena Zarowiecki, Malcolm Fisher, Christina James-Zorn, Virgilio Ponferrada, Aaron Zorn, Sridhar Ramachandran, Leyla Ruzicka, and Monte Westerfield. The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1):iyad031, May 2023. ISSN 1943-2631. doi: 10.1093/genetics/iyad031.
30. Sudhindra R. Gadagkar and Sudhir Kumar. Maximum Likelihood Outperforms Maximum Parsimony Even When Evolutionary Rates Are Heterotachous. *Molecular Biology and Evolution*, 22(11):2139–2141, November 2005. ISSN 0737-4038. doi: 10.1093/molbev/msi212.
31. Eric W. Sayers, Evan E. Bolton, J. Rodney Brister, Kathi Canese, Jessica Chan, Donald C. Comeau, Ryan Connor, Kathryn Funk, Chris Kelly, Sunghwan Kim, Tom Madej, Aron Marchler-Bauer, Christopher Lanczycki, Stacy Lathrop, Zhiyong Lu, Francoise Thibaud-Nissen, Terence Murphy, Lon Phan, Yuri Skripchenko, Tony Tse, Jiyao Wang, Rebecca Williams, Barton W. Trawick, Kim D. Pruitt, and Stephen T. Sherry. Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50(D1):D20–D26, January 2022. ISSN 1362-4962. doi: 10.1093/nar/gkab1112.
32. Gergely J. Szöllősi, Adrián Arellano Davín, Eric Tannier, Vincent Daubin, and Bastien Boussau. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 370 (1678):20140335, September 2015. ISSN 1471-2970. doi: 10.1098/rstb.2014.0335.
33. Aristotelis Tsirigos and Isidore Rigoutsos. A new computational method for the detection of horizontal gene transfer events. *Nucleic acids research*, 33(3):922–933, January 2005. ISSN 1362-4962. doi: 10.1093/nar/gki187.
34. Marek Kokot, Maciej Długosz, and Sebastian Deorowicz. KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics*, 33(17):2759–2761, September 2017. ISSN 1367-4803. doi: 10.1093/bioinformatics/btx304.
35. Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1):25–29, May 2000. ISSN 1546-1718. doi: 10.1038/75556.

Supplementary Information

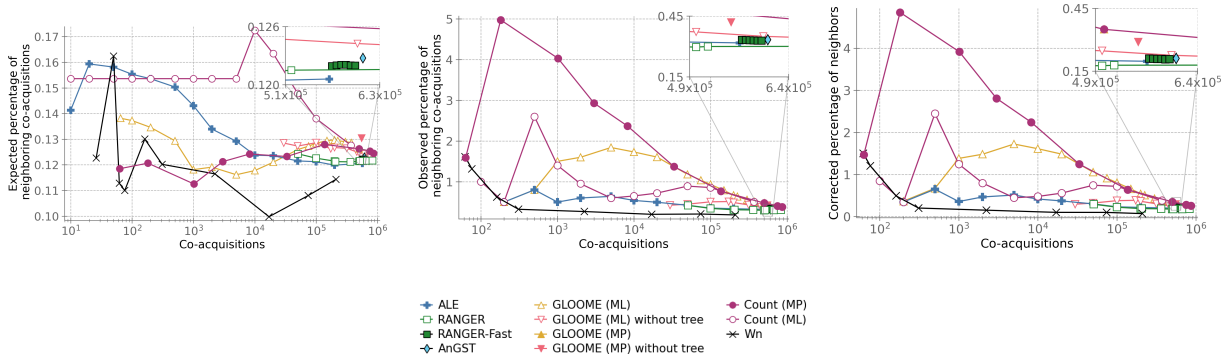

Fig. S1. (a) Expected, (b) observed, and (c) observed minus expected percentage of neighboring co-acquisitions, vs number of co-acquisitions inferred. The expected percentage of neighboring co-acquisitions changes with the number of co-acquisitions inferred at various thresholds, and varies between methods, since it depends on the set of chromosomes that those co-acquisitions are inferred on.

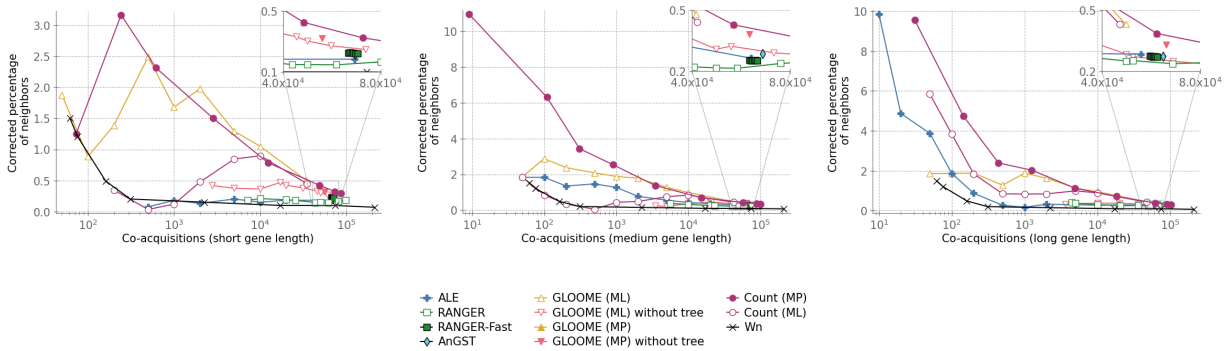

Fig. S2. Percentage of neighboring co-acquisitions inferred minus that expected by random chance, vs co-acquisitions, for gene families with mean gene length being (a) short (less than 625.58bp), (b) medium (between 625.58bp and 1028.54bp), and (c) long (greater than 1028.54bp), i.e. the lower, middle, and upper terciles of gene lengths.

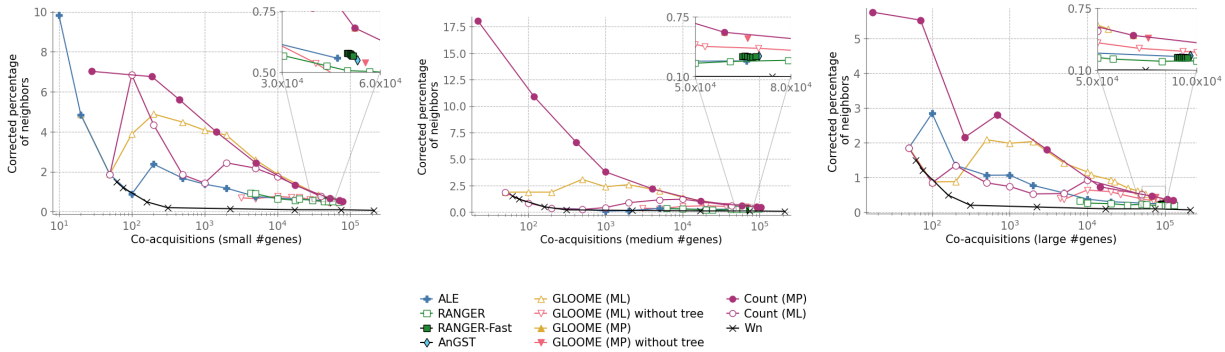

Fig. S3. Percentage of neighboring co-acquisitions inferred minus that expected by random chance, vs co-acquisitions, for gene families with number of genes being (a) small (less than 67), (b) medium (between 67 and 86) and (c) large (more than 86), i.e. the lower, middle, and upper terciles of gene family sizes.

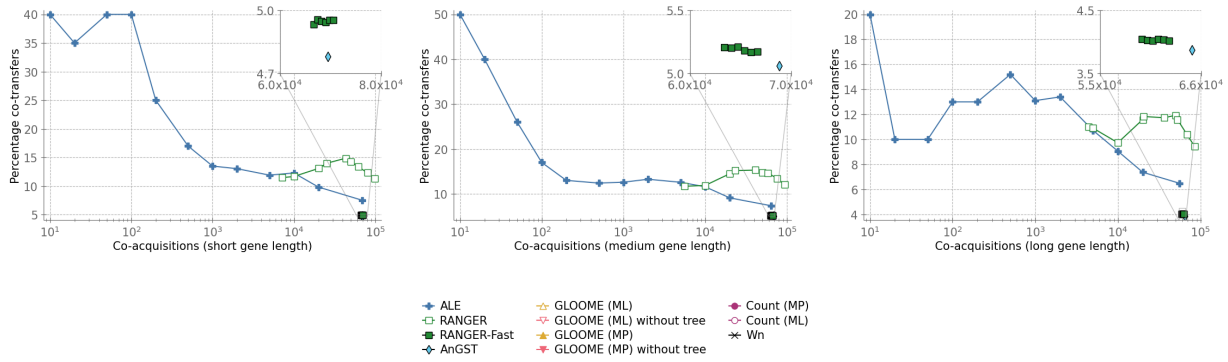

Fig. S4. Percentage of co-transfers vs co-acquisitions, for gene families with mean gene length being (a) short (less than 625.58bp), (b) medium (between 625.58bp and 1028.54bp), and (c) long (greater than 1028.54bp), i.e. the lower, middle, and upper terciles of gene lengths.

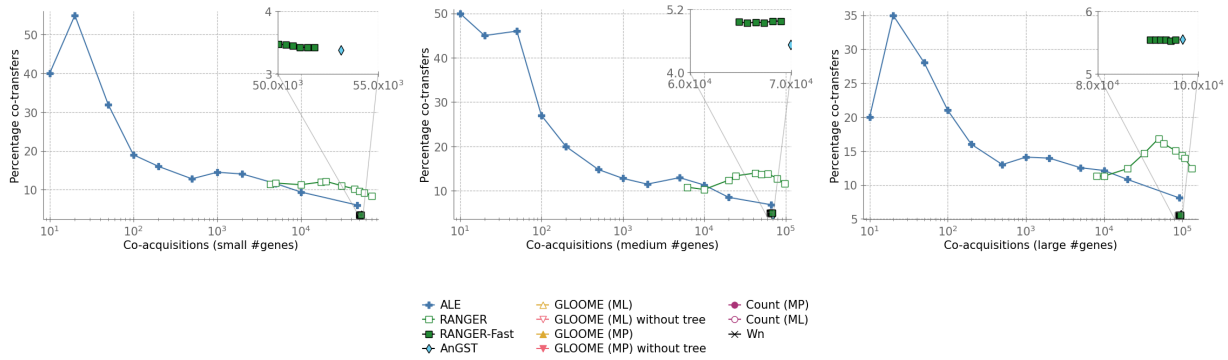

Fig. S5. Percentage of co-transfers vs co-acquisitions, for gene families with number of genes being (a) small (less than 67), (b) medium (between 67 and 86) and (c) large (more than 86), i.e. the lower, middle, and upper terciles of gene family sizes.

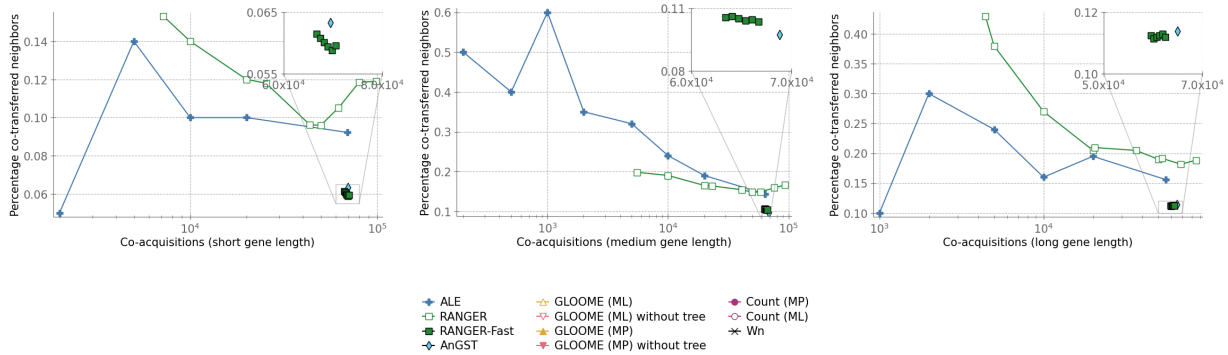

Fig. S6. Percentage of co-transferred neighbors vs co-acquisitions, for gene families with mean gene length being (a) short (less than 625.58bp), (b) medium (between 625.58bp and 1028.54bp), and (c) long (greater than 1028.54bp), i.e. the lower, middle, and upper terciles of gene lengths.

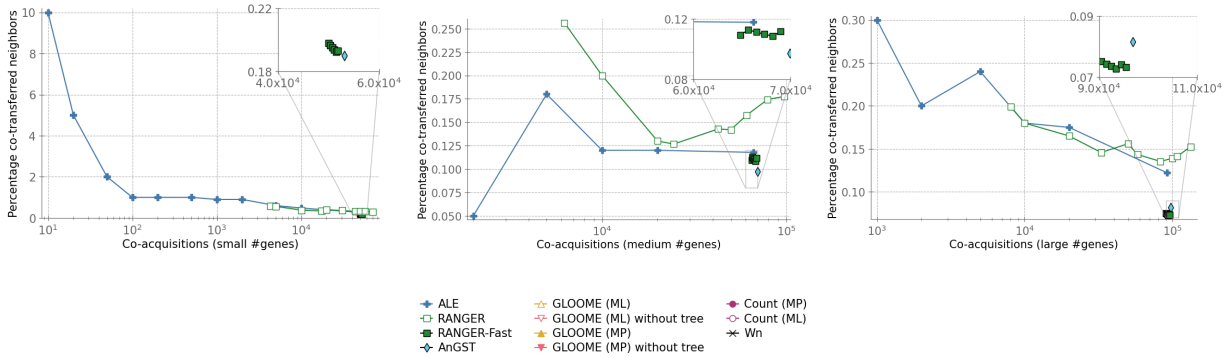

Fig. S7. Percentage of co-transferred neighbors vs co-acquisitions, for gene families with number of genes being (a) small (less than 67), (b) medium (between 67 and 86) and (c) large (more than 86), i.e. the lower, middle, and upper terciles of gene family sizes.

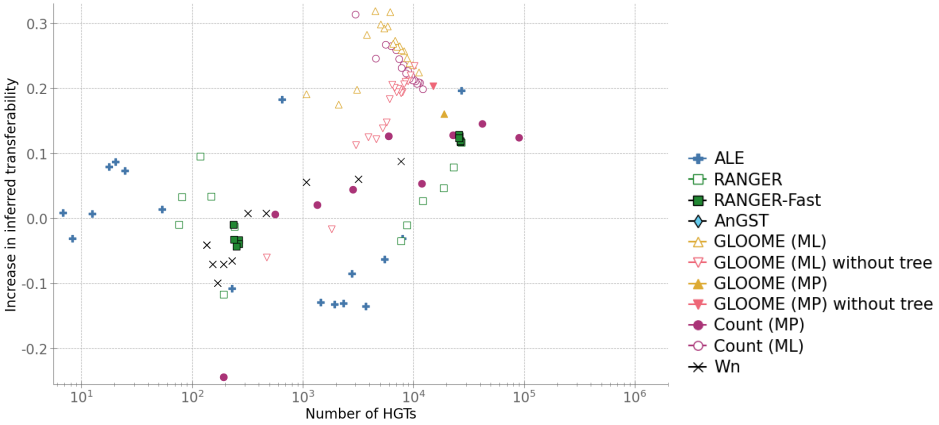

Fig. S8. Relative increase in number of transfers inferred for response genes, compared to non-response genes, for all methods at all transfer thresholds (including p-values >= 0.05).

Method	Mean (t=0)	σ (t=0)	Max (t=0)	Mean (t=2)	σ (t=2)	Max (t=2)	Mean (t=3)	σ (t=3)	Max (t=3)
Count (MP)	1.352	1.236	3.867	2.237	1.722	4.972	2.435	1.770	5.010
GLOOME (ML)	0.646	0.329	1.140	1.331	0.583	2.320	1.531	0.681	2.780
Count (ML)	0.425	0.331	1.200	1.178	0.743	3.200	1.499	0.868	3.800
ALE	0.391	0.191	0.800	0.629	0.178	1.000	0.734	0.200	1.040
Wn	0.353	0.546	1.613	0.705	0.578	1.613	0.778	0.528	1.613
GLOOME (MP)	0.271	0.000	0.271	0.637	0.000	0.637	0.778	0.000	0.778
GLOOME (ML)$^{\alpha}$	0.246	0.038	0.300	0.568	0.079	0.675	0.692	0.092	0.814
GLOOME (MP)$^{\alpha}$	0.242	0.000	0.242	0.563	0.000	0.563	0.687	0.000	0.687
AnGST	0.191	0.000	0.191	0.450	0.000	0.450	0.557	0.000	0.557
RANGER-Fast	0.190	0.001	0.192	0.440	0.001	0.442	0.543	0.001	0.545
RANGER	0.185	0.029	0.258	0.433	0.048	0.551	0.536	0.061	0.687

Table S1. Mean, standard deviation, and maximum of the corrected percentage of neighboring co-acquisitions inferred by each method, for $t \neq 1$, where t is the maximum number of intervening genes between two co-acquired pair of genes to be considered neighbors.
α : GLOOME without species tree.