Performance of a phylogenetic independent contrast method and an improved pairwise comparison under different scenarios of trait evolution after speciation and duplication

Tina Begum1,2 | Martha Liliana Serrano-Serrano1,2 | Marc Robinson-Rechavi1,2

Abstract

1. Despite the importance of gene function to evolutionary biology, the applicability of comparative methods to gene function is poorly known. A specific case which has crystallized methodological questions is the ‘orthologue conjecture’, the hypothesis that function evolves faster after duplication (i.e. in paralogues), and conversely is conserved between orthologues. Since the mode of functional evolution after duplication is not well known, we investigate under which reasonable evolutionary scenarios phylogenetic independent contrasts or pairwise comparisons can recover a putative signal of different functional evolution between orthologues and paralogues.

2. We investigate three different simulation models, which represent reasonable but simplified hypotheses about the evolution of a gene function trait. These are time-dependent trait acceleration, correlated changes in rates of both sequence and trait evolution and asymmetric trait jump. For each model we tested phylogenetic independent contrasts and an improved pairwise comparison method which accounts for interactions between events and node age.

3. Both approaches lose power to detect the trend of functional evolution when the functional trait accelerates for a long time following duplication for trees with many duplications, with better power of phylogenetic contrasts under intermediate scenarios. Concomitant increase in evolutionary rates of sequence and of trait after duplication can lead to both an incorrect rejection of the null under null simulations of trait evolution, and a false rejection of the orthologue conjecture under orthologue conjecture simulations, by phylogenetic independent contrasts. Improved pairwise comparisons are robust to this bias. Both approaches perform equally well under rapid shifts in traits.

4. Considering our ignorance of gene function evolution, and the potential for bias under simple models, we recommend methodological pluralism in studying gene family evolution. Functional phylogenomics is complex and results supported by only one method should be treated with caution.
1 | INTRODUCTION

In comparative biology, pairwise comparisons of terminal taxa of genes or of species are commonly used to detect phenotypic or morphological or character associations (Felsenstein, 1985; Maddison, 2000; Martins & Garland, 1991; Read & Nee, 1995). This pairwise approach can handle sparse data and poorly resolved phylogenies, and relies on relatively few assumptions, compared to other methods that use reconstructed ancestral states, stochastic models of evolution or phylogenetic branch lengths (Felsenstein, 1985; Garland et al., 1992; Maddison, 2000; Purvis & Bromham, 1997; Read & Nee, 1995). A potential issue is that multiple pairwise comparisons of data across the same branch of a tree pseudo-replicate data, and overlook phylogenetic relatedness of species (Blomberg et al., 2003; Dunn et al., 2018; Pagel, 1994). The phylogenetic independent contrasts (PIC) method solves these issues by taking into account the evolutionary history of species (Cooper et al., 2016; Dunn et al., 2018; Felsenstein, 1985; Grafen, 1989). This approach needs well-resolved phylogenies. Moreover, it is sensitive to model assumptions (e.g. known branch lengths of the phylogeny, or Brownian character evolution), and may lead to erroneous interpretations if these assumptions are violated (Cooper et al., 2016; Díaz-Uriarte & Garland, 1998; Freckleton & Harvey, 2006; Garland et al., 1992; Maddison, 2000). Most comparative functional genomic studies still rely on pairwise comparisons, although some do use PIC.

Applications of both approaches to test the ‘orthologue conjecture’ model, that is, that there are larger functional differences between paralogs (homologous genes diverging since a duplication) than between orthologs (homologous genes diverging since a speciation), have shown how methodological differences can lead to different results concerning the evolution of gene function (e.g. Begum & Robinson-Rechavi, 2021; Dunn et al., 2018; Kryuchkova-Mostacci & Robinson-Rechavi, 2016; Yanai et al., 2005). An advantage of simulations is that we have perfect knowledge of the original time trees (Figure 1a), which we never know for empirical data. With empirical data, we obtain ‘Substitution trees’ (Figure 1b), which are time calibrated using the speciation time points to generate pseudo time trees, that is, ‘Calibrated time trees’ (Figure 1c), for hypothesis testing. We compared the performances of PIC and of an improved pairwise method using three simple simulation models with different parameters of divergence following speciations and duplications. We used two sets of simulated trees: (a) trees with arbitrary parameter values and different proportions of duplications, which we call ‘pure simulated’ trees, and (b) trees simulated using parameter values (speciation rates, extinction rates, number of tips and proportions of internal node events) from calibrated dichotomous empirical gene trees. The first allows to explore parameter space in an unbiased way, while the latter allows to explore more realistic scenarios (Table S1) but can be affected by bias of duplication branch lengths (Begum & Robinson-Rechavi, 2021).

2 | MATERIALS AND METHODS

2.1 | Simulated gene trees

We simulated ultrametric trees (n = 10,000), each with 100 tips generated under a pure birth–death model using the TreeSim R package (Stadler, 2011) with a speciation rate of 0.4, and an extinction rate of...
0.1 for pure simulated trees (Table S1). Since our study aims to mimic simple evolutionary scenarios (test model details in Supplementary text), where standard trait evolutionary models are frequently applied on comparatively smaller phylogenies (Chira & Thomas, 2016), we limited our tree size to 100 tips. We call these simulated time trees ‘Original time trees’ (Figure 1a), with the branch lengths corresponding to units of time (i.e. Million Years – My). For consistency, we used the same set of 10,000 original time trees for all further analyses.

2.2 Annotation of internal node events

We annotated internal node events as ‘speciation’ or ‘duplication’ so that each pure simulated tree had at least one speciation and one duplication node events. We considered four proportions of duplication events (number of duplication events/total number of internal nodes): 0.1, 0.2, 0.5 and 0.8 (Figure S1; Table S1). We first randomly annotated ‘duplication’ events to internal nodes based on the proportion of duplication events we considered for the tree set. We then assigned ‘speciation’ events to the rest of the nodes of the tree. From empirical vertebrate data, we find the median proportion of duplications events in a range of 0.1–0.2 (Begum & Robinson-Rechavi, 2021; Dunn et al., 2018). Therefore, we used the pure simulated tree set with a proportion 0.2 of duplication events (Table S1) in all our main text analyses to compare the performances of different approaches in testing the orthologue conjecture on different simulation models. To verify that our results did not depend on the number of duplication events in a tree, we reproduced results using proportions of duplications of 0.1, 0.5 and 0.8 (supporting materials).

2.3 Simulation of trait

In this study, we considered a trait with a range between 0 and 1. We call it, \( r \) like the tissue-specificity score used in several previous studies of the orthologue conjecture (Dunn et al., 2018; Kryuchkova-Mostacci & Robinson-Rechavi, 2016; Yanai et al., 2005). For the sake of simplicity, we only used a Brownian model (BM) of trait evolution:

\[
\text{d}X(t) = \sigma^2 t.
\]

FIGURE 1 An example of simulated tree models used in this study. (a) Original (true) time tree, (b) substitution rates tree and (c) pseudo time tree inferred from (b), with simulated trait values at tips. We used a simulated tree with 20 tips for illustration. Branches following gene duplication are asymmetrically painted using the phyTools R package (Revell, 2012) to introduce different sequence evolutionary rates to those branches. This demonstrates how the scale (i.e. the branch lengths) of the calibrated tree can differ from the original time tree.
where $dX(t)$ quantifies the change in trait, and the Brownian variance $\sigma^2$ parameter describing the rate at which taxa diverge from each other through time $t$ (Cavalli-Sforza & Edwards, 1967; Chira & Thomas, 2016; Harmon et al., 2010). We used FastBM function of the PHTOOLS R package (Revell, 2012) to simulate the trait $r$ on each original time tree. The same trait values were used for the corresponding substitution rate trees (described later), and for the pseudo time calibrated trees (described later) for further analyses with phylogenetic method.

### 2.4 | Generation of calibrated pseudo time trees

We used the NELSI R package (Ho et al., 2015), and modified the original simulate.rates function to simulate.rates_heterogeneous to introduce different sequence evolutionary rates (nucleotide substitutions/site/My) to the branches of different node events with the aid of ‘discrete’ multi-rates clock model. To avoid bias in our inference, we introduced different sequence evolutionary rates ($K$, as used for simulation of the trait $r$ in a null scenario) to the branches of different node events with the aid of ‘discrete’ multi-rates clock model. To avoid bias in our inference, we used the same sequence evolutionary rates, $K$, as used for simulation of the trait $r$ in a null scenario. These trees are ‘Substitution trees’ (Figure 1b), where the branch lengths of each tree representing substitution rates. In these phylograms, the branch lengths are proportional to the number of mutations accumulated along the branches, and thus changes in branch lengths should have an influence on both the nonsynonymous and synonymous substitution rates.

We used the chronos function of the ape R package (Paradis et al., 2004) to generate calibrated time trees (Figure 1c). For such time calibration, we normally have access to only a few speciation time points (focal speciation nodes) for which we have external references (e.g. fossil data) (Begum & Robinson-Rechavi, 2021; Dunn et al., 2018). However, due to simulated tree structure, very few trees passed time calibration step when we fix focal speciation nodes, and their ages across all trees. Hence, we used the ages of all the speciation nodes of a tree to time calibrate them. The scale of the original tree drastically changes after calibration if we do not fix the age of the root node. It may impact the inference of phylogenetic method due to its branch length dependence. Hence, we also used the age of the root of each simulated original time tree to maintain the scale after time calibration.

### 2.5 | Simulations with empirical data parameters

We downloaded 60,447 gene trees from ENSEMBL Compara v.100 (Herrero et al., 2016), and annotated tips with precomputed tissue-specificity data, covering six organs of 21 vertebrate species from Fukushima and Pollock (2020). Tips with missing trait data were pruned. When two speciation nodes had the same clade names at different node depths of a tree, the older speciation event was edited to ‘NA’ to prevent failure of time calibration (Dunn et al., 2018). This led us to obtain 13,647 calibrated empirical gene trees.

To simulate original time trees based on empirical data parameters, we used only dichotomous trees, that is, 6,953 of 13,647. This helps us to avoid over-estimated speciation and extinction rates using the birthdeath function of the ape R package (Paradis et al., 2004). The estimated rate parameters for each tree, including tip number, were used to simulate 6,953 original time trees using TreeSim (Stadler, 2011). Since the actual tree topology changes after simulation, we randomly assigned node events for each tree, while maintaining the empirical event proportions (Table S1).

In this case, we estimated trait evolutionary rates for each calibrated time tree based on empirical $r$ of tips following a BM using the fitContinuous function (Pennell et al., 2014). We used them to simulate trait for our 3 test models.

### 2.6 | Phylogenetic analyses

For each internal node (inode) of a gene tree, we calculated the PIC of simulated trait $r$ from the ape R package using the following formula:

$$\text{PIC}_{\text{inode}} = \frac{r_1 - r_2}{\sqrt{\text{branchlength}_1 + \text{branchlength}_2}}.$$

where PIC$_{\text{inode}}$ is the node of which PIC is to be computed, $r_1$ and $r_2$ are the trait values of daughters of the corresponding node. The branch lengths of the daughters of corresponding node are represented by branchlength$_1$ and branchlength$_2$ respectively (Dunn et al., 2018; Felsenstein, 1985; Paradis et al., 2004). We then contrasted the PIC of each event (speciation or duplication) using a Wilcoxon two-tailed rank test.

### 2.7 | Pairwise analyses

We considered all possible pairwise comparisons for each tree to obtain Pearson correlation coefficients $r$ between $r$ values of events, as was done by Dunn et al. (2018). Few correlation points of speciation and duplication events often make it difficult to infer the statistical significance level precisely. Since most of the pure simulated trees have maximum node ages <20 My, we computed the correlation coefficient, $r$ of speciation and of duplication events over each 0.05 My time intervals. To analyse simulated trees with empirical parameters, the time interval is set to 10 My since we have maximum node ages >2,000 My. This gives the estimates of correlation for both the events at equal time points. Previous pairwise comparisons of Kryuchkova-Mostacci and Robinson-Rechavi (2016) used a linear model (i.e. lm(R ~ Event)) to distinguish the effects of events on correlation coefficients. In contrast, we used polynomial (linear, quadratic or cubic) regression in our improved pairwise analyses, and used repeated 10-fold cross-validation approach to choose the best-fit model. The model with the highest adjusted $R^2$ value and with the lowest root mean square error among the linear, quadratic or cubic models was chosen as the best-fit model for each case to avoid over fitting. We then compared with ANOVA the best-fit polynomial with and without the interaction terms between events and age, to obtain our improved pairwise $p$ values.
2.8 | Statistical significance

For any method, we considered a result significant if \( p < \alpha' \). We applied a Bonferroni correction over the 236 tests that we performed overall in this study, with \( \alpha = 0.05 \), thus \( \alpha' = 0.05/236 = 2.12 \times 10^{-6} \).

3 | RESULTS

3.1 | Model 1: Time-dependent trait acceleration model

Immediately after duplication, both duplicates might experience accelerated evolution for a period of time due to relaxed selection (Rogozin, 2014). This can be represented by a simple model of acceleration of trait evolutionary rate for a given time after duplication (Figure 2). After that period of acceleration, the evolutionary rates return to the pre-duplication level. In this model, speciation does not have any impact on evolutionary rates: a speciation during the time period of the duplication-caused acceleration does not revert the gene to a 'speciation' rate. To study our capacity to recover evolutionary signal under this model, we simulated with different durations of accelerated evolution, and different proportions of duplications to speciations in the gene trees (Figure 2). In this comparative study, we used an improved pairwise comparison due to the low power of the simple pairwise comparison approach used in Kryuchkova-Mostacci and Robinson-Rechavi (2016) (Tables S2 and S3).

First, both methods recover the proper evolutionary signal under the null, with no difference in PIC between duplication and speciation branches (Figure 3a), and no difference between orthologue and parologue pairs (Figure 3b). Second, when there are few duplications in proportion to speciations, both approaches recover the evolutionary signal, namely faster evolution of paralogues (pairwise) or of duplication branches (PIC) (Figure 3c–f; Tables S2 and S3). This is despite the acceleration lasting 50% to 90% of the time since the oldest duplication. This is possible because the rarity of duplications provides many speciation branches without acceleration (Figure 2). While the performance of both methods at recovering the evolutionary signal decreases when the proportion of duplication nodes increases, the PIC is more robust than the pairwise approach (Table S2). With 50% of duplication nodes, only the PIC still recovers the evolutionary trend when the trait acceleration period increases. At 80% of duplications and an acceleration lasting 70% to 90% of old duplication age, neither approach can recover the signal of the orthologue conjecture (Table S2). Under such scenarios, most speciation branches or orthologue pairs have evolved mostly or entirely under the influence of a duplication-caused acceleration, and there is no signal to recover.

We obtained similar trends for the trees simulated using empirical data parameters when we used sufficiently high trait acceleration rates following duplication (Table S4), although the PIC then failed under the null. With only a slight acceleration (1.2 or 2 times), the phylogenetic method incorrectly rejects the null under a null simulation, and shows an opposite trend under the orthologue conjecture simulations of trait evolution. In these cases, the improved pairwise comparisons can still recover simulated trends, in many scenarios including the null, although with low power (Table S4). The failure of PIC is probably due to the bias in branch lengths when using parameter values from calibrated empirical gene trees.

3.2 | Model 2: Rates of sequence and trait evolution model

While there is debate over the evolutionary rate of gene function after duplication, it has been shown repeatedly that there is an acceleration of sequence evolutionary rates (Brunet et al., 2006; Conant & Wagner, 2003; Cusack & Wolfe, 2007; Gu et al., 2005; Holland et al., 2017; Kim & Yi, 2006; Panchin et al., 2010; Pegueroles et al., 2013; Pich i Roselló & Kondrashov, 2014; Scannell & Wolfe, 2008; Studer & Robinson-Rechavi, 2009). We explore the impact of this acceleration on the PIC and improved pairwise comparisons under different scenarios (Table 1). Because many empirical studies have found asymmetry of this acceleration of duplicate gene sequence evolution, we used an asymmetric acceleration in our simulations.

Unlike under the previous model, the proportion of duplication events in the trees has no impact on the capacity to recover trends (Figure 4, Figures S2–S5). An important difference on the other hand is that the null is not always properly recovered. When the trait evolutionary rate is not impacted by duplication, but the sequence evolutionary rate is (scenario 2 of Table 1), the PIC approach wrongly rejects the null for trait evolution. This is due to time tree calibration using sequence evolution as a reference. Indeed, the PIC does not reject the null when the original time tree is used (Figure A in S2, C in Figures S3–S7). However, the higher sequence evolutionary rates after duplication lead to higher expected variance, and thus the calibrated time trees produce lower contrasts for duplication than for speciations, with a significant difference in PIC (Figure A in S2, C in Figures S3–S7). Moreover, if both the trait and the sequence evolutionary rates accelerate after duplication events, but the sequence acceleration is higher (scenario 4c of Table 1), the PIC can again estimate lower contrasts for duplication branches with calibrated time trees (Figure 4e, K in Figures S3–S7). This can lead to wrongly accepting the null if a one-sided test is used, or rejecting the null in the wrong direction. This bias does not impact the pairwise approach, under any of these scenarios (Figure 4f; B in Figure S2, D and L in Figure S3–S7). Thus, even though the assumption of independent data points is violated, the improved pairwise approach is more robust than the PIC in this type of model, since it does not depend on the branch lengths.
3.3 | Model 3: Asymmetric trait jump model

Finally, we investigated our ability to recover gene evolutionary patterns under a jump model rather than changes in continuous rates. The idea is that, similarly to rapid trait jumps when organisms transition into new adaptive zones (Duchen et al., 2017; Simpson, 1944), paralogues might undergo rapid trait jumps when they appear, for example, because of a new chromosomal environment in the case of asymmetric duplication. We used a Brownian motion model with asymmetric jumps in trait (Figure 5). Jumps can follow either speciation or duplication events, and different gene evolution models correspond to different probabilities of jumps after the two types of events. We used a simple model where there can only be one or zero jump per branch. Speciation and duplication branches are randomly
chosen, and one of their daughter branches (randomly selected) experiences a rapid jump in trait, then returns to progressive trait evolution along the branch; there is only one progressive evolutionary rate. Under the null, equal proportions of speciation and duplication branches (30%) were randomly chosen. Reassuringly, the null was rejected neither by PIC (A in Figure S8) nor by the pairwise approach (B in Figure S8) for the standard pure simulation models. For the orthologue conjecture, we simulated more asymmetric jumps following duplications than speciations. More jumps in trait should introduce more divergence in trait, and this should be detected as support for the orthologue conjecture. Indeed, both approaches did support the orthologue conjecture in this case (C and D in Figure S8). While the most naive form of this orthologue conjecture is one evolutionary rate on duplication branches, and another on speciation branches, we explored three more complex models of trait evolution.

It is worth mentioning that the models under which we have simulated cannot be easily tested on empirical data, since we lack information about the original time trees on which traits have evolved. Moreover, on empirical trees it is impossible to guarantee a true null model including no change in sequence evolutionary rates after duplication. We rather obtain calibrated pseudo time trees in reality. Such calibrated empirical gene trees cannot always be considered

4 | DISCUSSION

A good method to detect patterns of gene functional evolution should not reject the null hypothesis when data are simulated under the null, should successfully reject the null when simulated under an alternative hypothesis, and should recover the right direction of pattern difference. Here, our null hypothesis is that gene function evolves independently of duplication, and our alternative is the ‘orthologue conjecture’, that function evolves faster after duplication. While the most naive form of this orthologue conjecture is one evolutionary rate on duplication branches, and another on speciation branches, we explored three more complex models of trait evolution.

It is worth mentioning that the models under which we have simulated cannot be easily tested on empirical data, since we lack information about the original time trees on which traits have evolved. More
as good proxies of the original time trees, since the branch lengths of such trees are biased, especially for old duplicates, due to the non-availability of reference time points for duplications (Begum & Robinson-Rechavi, 2021). When traits are simulated based on such biased calibrated time trees, the performances of both approaches deteriorate by recovering opposite pseudo-signals in the null condition (Begum & Robinson-Rechavi, 2021; Dunn et al., 2018). Thus they may not be sufficient to trace expected patterns of trait evolution in many realistic scenarios. Even with trees simulated using parameter values from such calibrated time trees, failure to recover the expected trends with sufficient power is common. This is due to the use of biased phylogenies for parameter estimation for simulations. On the contrary, pure simulation models are devoid of such calibration bias, and thus are suitable to investigate the behaviour of methods in theoretical scenarios where testing with empirical data seems impossible, including the pure null. If we could obtain calibrated unbiased empirical phylogenies, the trends of functional evolution should be consistent with these pure simulation models for the testable cases, since the pure simulated trees with different proportions of duplications allow to understand the patterns of functional evolution irrespective of the positions of events in trees. We will thus focus on the result from the pure simulated trees (Table 2).

First, while the jump model is very different from the most naive orthologue conjecture model, both methods tested (pairwise and PIC) performed well under it unless affected by biased phylogenetic patterns (Figures S8–S13). Thus, testing the orthologue conjecture seems robust to change being sudden or gradual. Second, when gradual change extends in time to both speciation and duplication branches, methods lose power to detect the correct pattern. This loss increases with the duration of acceleration and with the proportion of branches affected. This time-dependent trait acceleration model represents a reasonable assumption of gene evolution, as there is no strong reason to expect paralogues to stop evolving like paralogues because there was a speciation. Indeed duplication affects each gene’s direct environment, whereas speciation can have at most a very indirect effect. The PIC was more powerful than the pairwise approach under this model, on pure simulated trees (Tables S2 and S3). Third and most worryingly, when both sequence and trait evolutionary rates are affected by duplication, PIC can both
incorrectly reject the null under null simulations of trait evolution, and report the opposite direction of change under orthologue conjecture simulations (Figure 4e, A in S2, C and K in S3–S7 Figures).

While all of the simulation models used are quite simple, they represent realistic scenarios of gene function evolution. The mechanisms of trait divergence following gene duplication are still poorly understood. For example, we do not have direct evidence for an asymmetric trait jump model or for a time-dependent trait acceleration model. The purpose of our simulation study was to identify scenarios where the PIC and the pairwise comparisons may produce
distinct signals of functional evolution between orthologues and paralogues. Although our simple simulations used the same mechanism of trait evolution for all the duplicates of a tree, a mix of different models for duplicates in different parts or at different times in a same gene tree cannot be excluded.

Pairwise comparisons of traits between genes have been criticized in studying functional evolution (Dunn et al., 2018). Behaviour under a null model can help to assess the presence of a systematic bias in the approach. In all different models, and with different sets of trees (Table S1), we found that our improved pairwise comparisons gave the expected results under the null. The simple pairwise comparison of Kryuchkova-Mostacci and Robinson-Rechavi (2016) also produced the expected signal under the null (Tables S2 and S3), although it lacked power under the alternative model. The largest difference between pairwise and PIC was the failure of PIC when both rates of sequence and of trait are affected by duplication (Figure 4; Figures S2–S7). This is worrying since acceleration of sequence evolutionary rates is common following gene duplications (Assis & Bachtrog, 2013; Gu et al., 2005; Jiang & Assis, 2017; Kryuchkova-Mostacci & Robinson-Rechavi, 2016; Lafond et al., 2018; Panchin et al., 2010; Pegueroles et al., 2013; Scannell & Wolfe, 2008). Thus, the scenario of these simulations is expected to be widespread in real data. In this case, higher sequence evolutionary rates for duplicates lead to higher expected variances, and thus produce lower PIC for duplication events than speciation events. These lower PIC are in contrast to the null expectation of no difference.

Our simulations also show that in the most probable orthologue conjecture scenario, where both the trait and the
sequence evolutionary rates accelerate after duplication events (Gu et al., 2005), but the ratio of sequence to the trait acceleration is higher (scenario 4c of Table 1), the PIC method fails to detect the signal of the orthologue conjecture (Figure 4e, K in S3–S7 Figures). In this case, improved pairwise comparisons seem to be a better choice due to its branch length independence.

The only model for which the ratio of duplication to speciation events impacts the results is with time-dependent trait acceleration. When there are many duplications, and the duplication-induced acceleration lasts for 70% to 90% of old duplication age, most speciation branches also evolve under the influence of trait acceleration like duplicates. Hence, both the approaches can fail to recover the signal of the orthologue conjecture in a real orthologue conjecture scenario (Table S2). Under intermediate scenarios, PIC allows to better recover the signal of the orthologue conjecture than pairwise comparisons. However, such a pattern is rare in the empirical parameter space we used, since they have low proportions of duplications, and both the approaches performed well under it (Tables S2 and S3).

Our study uses a BM that accounts only for the drift parameter. It is thus easier to infer the results of such a model than that of an Ornstein–Uhlenbeck (OU) model, which includes an additional \( \alpha \) parameter to account for selection. It should be noted that most phylogenetic methods, notably the PIC, were developed under a pure neutral model of evolution, that is, under a BM. This means that the inference of the results of a PIC from an OU model is not so straightforward, and we need special care before we draw an inference from the PIC on OU (Begum & Robinson- Rechavi, 2021). Thus, using OU could add another level of complexity in the inference of our complex evolutionary models of gene duplications. However, such application of BM is not irrelevant since many duplicates are likely to undergo selection in a genome, but not all. It has been shown before that duplicates, especially from whole genome duplication, can be retained in the genome due to dosage constraints (Gout & Lynch, 2015; Thompson et al., 2016). Moreover, a BM has already been applied to modelling gene duplication (e.g. Gu et al., 2005).

Overall, the main message of this study is that our ignorance of gene function evolution makes it difficult to choose one method as the gold standard. Each method has its own limitations, and we call for methodological pluralism, at least in the present state of our knowledge. Relying on a single approach to interpret result can be problematic in many cases of comparative functional genomics study involving complex gene family evolution. These results apply in principle not only to the orthologue conjecture, but also to any other cases of gene trait evolution (e.g. horizontal gene transfer), where evolution might be gradual or by jumps, affect a more or less large proportion of branches of the gene tree, and be confounded with changes in sequence evolutionary rates. In conclusion, functional phylogenomics is complex and results supported by only one method should be treated with caution.

ACKNOWLEDGEMENTS
We sincerely thank Julien Wollbrett, Elsa Guillot, Sara Fonseca Costa and all the members of the Robinson-Rechavi and Salamin groups for their help and useful discussions. Parts of the computations were performed at the Vital-IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics, as well as using the Wally cluster of the University of Lausanne. Work supported by Swiss National Science Foundation grant 31003A_173048.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORS’ CONTRIBUTIONS
T.B., M.L.S.-S. and M.R.-R. conceived the ideas; T.B. and M.L.S.-S. designed the original methodology; T.B. and M.R.-R. refined the methodology and led the writing of the manuscript; T.B. and M.L.S.-S. wrote the code for analysis with input from M.R.-R.; T.B. analysed the data. All authors contributed critically to the drafts and gave final approval for publication.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13680.

DATA AVAILABILITY STATEMENT
For reproducibility, data are archived at Zenodo http://doi.org/10.5281/zenodo.5091741 (Begum et al., 2021). All scripts are available at https://github.com/tbegum/Method-matters-in-gene-family-evolution.

ORCID
Marc Robinson-Rechavi https://orcid.org/0000-0002-3437-3329

REFERENCES
Assis, R., & Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in drosophila. Proceedings of the National Academy of Sciences of the United States of America, 110, 17409–17414. https://doi.org/10.1073/pnas.1313759110
Begum, T., & Robinson-Rechavi, M. (2021). Special care is needed in applying phylogenetic comparative methods to gene trees with speciation and duplication nodes. Molecular Biology and Evolution, 34, 1614–1626. https://doi.org/10.1093/molbev/msaa288
Begum, T., Serrano-Serrano, M. L., & Robinson-Rechavi, M. (2021). Method matters in gene family evolution. Zenodo, https://doi.org/10.5281/zenodo.5091741
Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution, 57, 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x
Bokma, F. (2008). Detection of ‘punctuated equilibrium’ by bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. Evolution, 62, 2718–2726. https://doi.org/10.1111/j.1558-5646.2008.00492.x
Brunet, F. G., Roest Crollius, H., Paris, M., Aury, J. M., Gibert, P., Jaillon, O., Laudet, V., & Robinson-Rechavi, M. (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Molecular Biology and Evolution, 23, 1808–1816. https://doi.org/10.1093/molbev/msl049
Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis. Models and estimation procedures. American Journal of Human Genetics, 19, 233–257.
Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rogozin, I. B. (2014). Complexity of gene expression evolution after duplication: Protein dosage rebalancing. *Genetics Research International*, 2014, 1–8. https://doi.org/10.1155/2014/516508

Scannell, D. R., & Wolfe, K. H. (2008). A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Research*, 18, 137–147. https://doi.org/10.1101/gr.6341207

Simpson, G. (1944). *Tempo and mode in evolution. A wartime book*. Columbia University Press.

Stadler, T. (2011). Simulating trees with a fixed number of extant species. *Systematic Biology*, 60, 676–684. https://doi.org/10.1093/sysbio/syr029

Stamboulian, M., Guerrero, R. F., Hahn, M. W., & Radivojac, P. (2020). The ortholog conjecture revisited: The value of orthologs and paralogs in function prediction. *Bioinformatics*, 36, i219–i226. https://doi.org/10.1093/bioinformatics/btaa468

Studer, R. A., & Robinson-Rechavi, M. (2009). How confident can we be that orthologs are similar, but paralogs differ? *Trends in Genetics*, 25, 210–216. https://doi.org/10.1016/j.tig.2009.03.004

Thompson, A., Zakon, H. H., & Kirkpatrick, M. (2016). Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. *Genetics*, 202, 765–774. https://doi.org/10.1534/genetics.115.178137

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domanay, E., Lancet, D., & Shmueli, O. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, 21, 650–659. https://doi.org/10.1093/bioinformatics/bti042

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article**: Begum, T., Serrano-Serrano, M. L., & Robinson-Rechavi, M. (2021). Performance of a phylogenetic independent contrast method and an improved pairwise comparison under different scenarios of trait evolution after speciation and duplication. *Methods in Ecology and Evolution*, 12, 1875–1887. https://doi.org/10.1111/2041-210X.13680