Quantifying Phylogenetic Beta Diversity: Distinguishing between 'True' Turnover of Lineages and Phylogenetic Diversity Gradients

Fabien Leprieur[1]*, Camille Albouy[1], Julien De Bortoli[1], Peter F. Cowman[2], David R. Bellwood[2], David Mouillot[1,2]

1 Laboratoire Ecologie des Systèmes Marins Côtiers UMR 5119, Université Montpellier 2, Montpellier, France, 2 Australian Research Council Centre of Excellence for Coral Reef Studies and School of Marine and Tropical Biology, James Cook University, Townsville, Queensland, Australia

Abstract

The evolutionary dissimilarity between communities (phylogenetic beta diversity PBD) has been increasingly explored by ecologists and biogeographers to assess the relative roles of ecological and evolutionary processes in structuring natural communities. Among PBD measures, the PhyloSor and UniFrac indices have been widely used to assess the level of turnover of lineages over geographical and environmental gradients. However, these indices can be considered as 'broad-sense' measures of phylogenetic turnover as they incorporate different aspects of differences in evolutionary history between communities that may be attributable to phylogenetic diversity gradients. In the present study, we extend an additive partitioning framework proposed for compositional beta diversity to PBD. Specifically, we decomposed the PhyloSor and UniFrac indices into two separate components accounting for 'true' phylogenetic turnover and phylogenetic diversity gradients, respectively. We illustrated the relevance of this framework using simple theoretical and archetypal examples, as well as an empirical study based on coral reef fish communities. Overall, our results suggest that using PhyloSor and UniFrac may greatly over-estimate the level of spatial turnover of lineages if the two compared communities show contrasting levels of phylogenetic diversity. We therefore recommend that future studies use the 'true' phylogenetic turnover component of these indices when the studied communities encompass a large phylogenetic diversity gradient.

Introduction

Phylogenies are increasingly used (i) to understand the origins and histories of species within a community (i.e. alpha diversity), (ii) to assess the relative roles of environmental sorting, competitive exclusion and evolutionary and biogeographical processes in shaping community structure [1–3] (iii) to predict the level of ecosystem functioning [4] and the delivery of services [5], and (iv) to guide conservation prioritization [6,7]. These arguments have recently been extended to phylogenetic beta diversity (PBD hereafter) that measures the phylogenetic dissimilarity among communities [8–10]. These authors argue that PBD allows a better understanding of the mechanisms underlying current biodiversity patterns by connecting local processes (e.g. biotic interactions and environmental filtering) with more regional processes, including trait evolution, speciation and dispersal.

To quantify PBD, numerous measures have been proposed [11], and more particularly two indices that derive from the taxonomic-based Sorenson and Jaccard's dissimilarity indices, namely the PhyloSor [8] and UniFrac [12] indices. These two closely related indices belong to a family of phylogenetic diversity-based dissimilarity measures, i.e. based on calculations using branch lengths [13]. For instance, the UniFrac index is expressed

as the total branch length unique to each community relative to the total branch length linking all species in both communities and hence measures the proportion of evolutionary history unique to each community [12].

Numerous studies have employed the PhyloSor and UniFrac indices to explore the spatial turnover of lineages over large spatial scales [3,10,13–15]. However, these indices are fundamentally based on 'broad-sense' measures of compositional beta diversity (CBD) that do not adjust for differences in composition attributable to richness gradients [16]. Consequently, the compositional differences arising from differences in species richness (species loss or gain associated with nestedness) cannot be distinguished from differences in species composition that are independent of species richness ('true' species turnover that involves species replacement) [17–19]. For instance, recent studies showed that using 'broad-sense' measures of CBD could make it difficult to tease apart the relative roles of neutral vs. niche-based processes in shaping CBD patterns [20,21]. This is because spatial turnover of species and nestedness of assemblages are two antithetic phenomena that are caused by different processes [18,20]. In that context, Baselga [19] proposed an additive partitioning framework that consists in decomposing the pair-wise Sørensen's dissimilarity index into two additive components

accounting for (i) 'true' turnover of species and (ii) richness differences between nested communities. Recent studies distinguishing between the 'true' turnover and nestedness components of CBD provided new insights into the mechanisms that drive CBD at large spatial scales [20–24].

As for their taxonomic-based relatives, the PhyloSor and UniFrac indices can be both considered as 'broad-sense' measures of phylogenetic turnover (i.e. incorporating differences in evolutionary history between communities attributable to phylogenetic diversity gradients). Recently, Ives & Helmus [25] tackled this issue by proposing a PBD metric that is independent of species richness in communities. However, this PBD metric does not control for differences in phylogenetic diversity between communities.

In the present study, we therefore extended the framework proposed by Baselga [19,26] to PBD by decomposing the PhyloSor and UniFrac indices into two components accounting for 'true' phylogenetic turnover and phylogenetic diversity gradients, respectively. To illustrate this decomposition, we used theoretical and archetypal examples. We also generated a large number of simulated communities from two contrasted types of phylogenetic tree, i.e. using either PDA (proportional-to-distinguishable arrangements) or Yule model. Finally, we used coral reef fish as a biological model to exemplify the relevance of distinguishing between 'true' phylogenetic turnover and phylogenetic diversity gradients when analysing large-scale patterns of PBD with marked differences in species richness and phylogenetic diversity.

Distinguishing between Phylogenetic Diversity Gradients and Spatial Turnover of Lineages: Formulations

Using an additive partitioning framework, Baselga [19] provided two separate components of species turnover and nestedness underlying the total amount of CBD. Specifically, this framework consists in decomposing the pair-wise Sørensen dissimilarity index (β_{sor}) into two additive components accounting for pure species turnover (β_{sim}) and nestedness (β_{sne}) patterns. The Simpson's dissimilarity index (β_{sim}) describes species turnover (or species replacement) without the influence of richness gradients [16,27–29]. Using basic operations on fractions, Baselga [19] derived a Nestedness-resultant dissimilarity index (β_{sne}) and showed that β_{sne} is simply the difference between β_{sor} and β_{sim} (i.e. $\beta_{sor} = \beta_{sim}+\beta_{sne}$). Specifically, β_{sne} reflects the increasing dissimilarity between nested assemblages due to the increasing differences in species richness [19]. These pairwise dissimilarity indices are formulated as:

$$\beta_{sor} = \frac{b+c}{2a+b+c} \quad (1)$$

$$\beta_{sim} = \frac{\min(b,c)}{a+\min(b,c)} \quad (2)$$

$$\beta_{sne} = \frac{\max(b,c)-\min(b,c)}{2a+b+c} \times \frac{a}{a+\min(b,c)} \quad (3)$$

where a is the number of species common to both sites, b is the number of species that occur in the first site but not in the second and c is the number of species that occur in the second site but not in the first [19].

Recently, Baselga [26] proposed a similar decomposition based on the Jaccard's dissimilarity index. The following pairwise dissimilarity indices (formulas 5 and 6) represent, the turnover and nestedness components of the Jaccard's dissimilarity index ($\beta_{jac} = \beta_{jtu}+\beta_{jne}$), respectively.

$$\beta_{jac} = \frac{b+c}{a+b+c} \quad (4)$$

$$\beta_{sim} = \frac{\min(b,c)}{a+\min(b,c)} \quad (5)$$

$$\beta_{jne} = \frac{\max(b,c)-\min(b,c)}{a+b+c} \times \frac{a}{a+2\min(b,c)} \quad (6)$$

Specifically, β_{jtu} measures the proportion of species that would be replaced between communities if both communities had the same number of species, and hence accounts for species replacement without the influence of richness difference [26]. In contrast, β_{jne} reflects the increasing dissimilarity between nested assemblages due to the increasing differences in species richness. Baselga [26] showed that the results obtained by the closely related Jaccard and Sørensen's dissimilarity indices were roughly equivalent.

Considering two communities j and k for which biodiversity can be quantified in terms of phylogenetic trees (T_j and T_k are the subset of a rooted regional tree T), we can express a as the sum of lengths for branches that are shared between communities j and k, b as the sum of lengths for branches that are present in community j but not found in assemblage k, c as the sum of lengths for branches that are present in community k but not found in community j. We express b, c and a using the phylogenetic diversity index [30,31] that can be calculated as the total branch length of a phylogenetic tree T that contains all species present in a community. Each branch t in the tree T has a length of w_t.

$$PD = \sum_T w_t \quad (7)$$

$$PD_{tot} = \sum_{T_j \cup T_k} w_t \quad (8)$$

$$PD_k = \sum_{T_k} w_t \quad (9)$$

$$PD_j = \sum_{T_j} w_j \quad (10)$$

$$b = PD_{Tot} - PD_k \quad (11)$$

$$c = PD_{Tot} - PD_j \quad (12)$$

$$a = PD_k + PD_j - PD_{Tot} \quad (13)$$

PhyloSor

$$= \frac{PD_{Tot} - PD_j + PD_{Tot} - PD_k}{2(PD_j + PD_k - PD_{Tot}) + PD_{Tot} - PD_j + PD_{Tot} - PD_k} \quad (14)$$

$$= \frac{2PD_{Tot} - PD_k - PD_j}{PD_k + PD_j}$$

UniFrac

$$= \frac{PD_{Tot} - PD_j + PD_{Tot} - PD_k}{PD_j + PD_k - PD_{Tot} + PD_{Tot} - PD_j + PD_{Tot} - PD_k} \quad (15)$$

$$= \frac{2PD_{Tot} - PD_k - PD_j}{PD_{Tot}}$$

Overall, both the PhyloSor and UniFrac indices range from 0 (the two communities are composed of similar species and hence share the same branches in the rooted phylogenetic tree) to 1 (the two communities are composed of distinct species that share no branch in the rooted phylogenetic tree). The two indices differ only because PhyloSor double weights the branch lengths shared by the two communities (i.e. the denominator of PhyloSor corresponds to the sum of phylogenetic diversity characterizing each community).

Following the formula (2), we obtained the turnover components of the PhyloSor and UniFrac indices, i.e. PhyloSor$_{Turn}$ and UniFrac$_{Turn}$, respectively:

$$PhyloSor_{Turn} =$$

$$\frac{\min\left(PD_{Tot} - PD_k, PD_{Tot-PD_j}\right)}{PD_k + PD_j - PD_{Tot} + \min\left(PD_{Tot} - PD_k, PD_{Tot} - PD_j\right)} \quad (16)$$

$$UniFrac_{Turn} =$$

$$\frac{2\min(PD_{Tot} - PD_k, PD_{Tot} - PD_j)}{PD_k + PD_j - PD_{Tot} + 2\min(PD_{Tot} - PD_k, PD_{Tot} - PD_j)} \quad (17)$$

The phylogenetic diversity (hereafter PD) component of PhyloSor is simply the difference between PhyloSor and PhyloSor$_{Turn}$, i.e. PhyloSor$_{PD}$ = PhyloSor – PhyloSor$_{Turn}$. It can be expressed using the formula (3) and by replacing a, b and c by the formula (10), (9) and (8), respectively. Similarly, the PD component of UniFrac is the difference between UniFrac and UniFrac$_{Turn}$, i.e. UniFrac$_{PD}$ = UniFrac – UniFrac$_{Turn}$. It can be expressed using the formula (6) and by replacing a, b and c by the formula (10), (9) and (8), respectively. All the analyses presented in this study were performed using the R statistical and programming environment [32]. The R code required to apply the additive partitioning framework is provided as Supporting Information (File S1 and File S2), together with the community dataset and the phylogenetic tree used to exemplify our approach (File S3 and File S4, respectively).

As a simple illustration of the proposed decomposition of PBD, the figure 1 presents three different examples. The first two examples (Fig. 1a and 1b, respectively) show two communities (A and B) that have no species in common ($\beta_{sor} = \beta_{sim} = 1$ and $\beta_{sne} = 0$). However, communities A and B (example 1, Fig. 1a.) display distantly related species, hence indicating locally phylogenetically clustered communities that have high PBD (PhyloSor = 1, i.e. the two communities compared do not share evolutionary
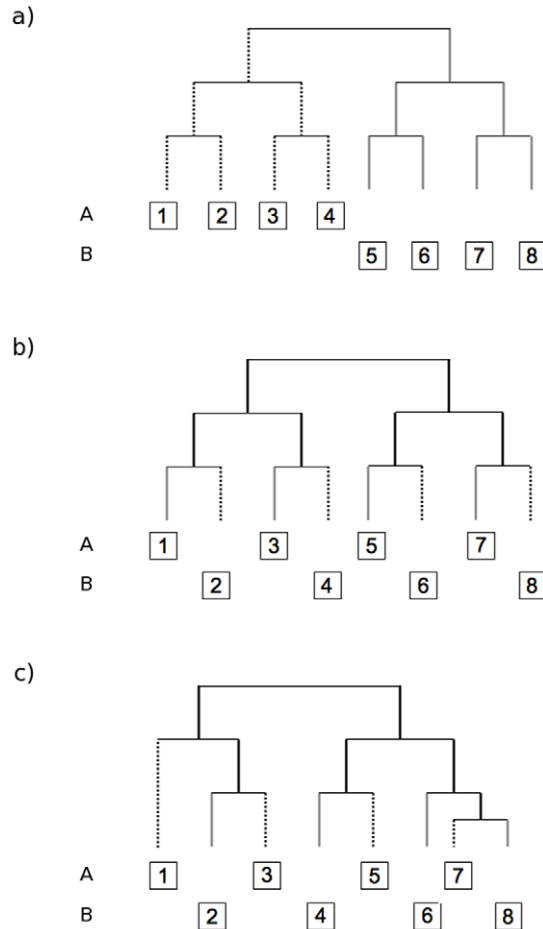

Figure 1. Three simple examples of phylogenetic tree associated to a pair of communities. Each community is composed of four species. All species are scored as present or absent in each example. All branch lengths are set to one, except for the last example (c) for which species 7 and species 8 display a branch length equal to 0.5. The three examples show similar level of compositional beta diversity (i.e. communities have no species in common) but differ in regards to the level of phylogenetic beta diversity (see main text for more details).
doi:10.1371/journal.pone.0042760.g001

history). In contrast, communities A and B illustrating the example 2 (Fig. 1b) display closely related species, hence indicating locally phylogenetically overdispersed communities that have little PBD (PhyloSor = 0.4, i.e. the two communities share a large amount of evolutionary history). For both examples, PhyloSor = PhySor$_{Turn}$ as PhySor$_{PD}$ = 0. For the first example, the PD component of PBD is zero because the two communities do not share any branch length and also display a similar level of PD (PD$_A$ = PD$_B$ = 7, see Fig. 1a). For the second example, the PD component of PBD is zero only because the two communities compared display the same level of PD (PD$_A$ = PD$_B$ = 10, see Fig. 1b). Indeed, if we reconsider the example 2 with two communities having slightly unequal levels of PD (Fig. 1c, PD$_A$ = 10 and PD$_B$ = 9), the overall level of PBD (PhyloSor = 0.421) is found to be different from the turnover component of PBD (PhySor$_{Turn}$ = 0.388). The difference between the two indices (expressed as PhySor$_{PD}$) quantifies the amount of PBD caused by a difference in PD between the two communities.

Let now consider 6 different communities sharing only two species (species 2 and 3 in Fig. 2). The phylogenetic diversity
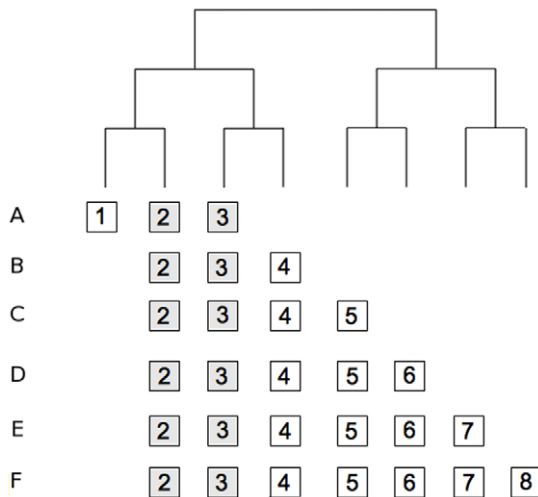
Figure 2. An example of six communities associated to a hypothetical phylogenetic tree. All species are scored as present or absent in each example and all branch lengths are set to one. Phylogenetic beta diversity (PBD) values were computed for several pairs of communities according to the PhyloSor and UniFrac indices and their respective turnover and phylogenetic diversity components (see Table 1 and main text for more details).
doi:10.1371/journal.pone.0042760.g002

unique to community A remains constant while the phylogenetic diversity unique to the other communities increases from B to F. Comparisons between the community A and the other communities (B to F, see Table 1) show that the increasing phylogenetic beta diversity (PhyloSor and UniFrac) is entirely caused by an increasing contribution of the PD component (PhySor$_{PD}$ and UniFrac$_{PD}$), while the turnover component remains constant across comparisons (PhySor$_{Turn}$ = 0.166 and UniFrac$_{Turn}$ = 0.286).

It is worth noting that PhyloSor$_{PD}$ = PhyloSor (or UniFrac$_{PD}$ = UniFrac) when the two communities compared are completely nested in regards to their taxonomic composition (see Fig. 2 and Table 1). For instance, the community B has no unique species and hence the branch length unique to community B is zero. Comparisons between the community B and the other communities (C to F, see Table 1) show that the increasing phylogenetic beta diversity (PhyloSor and UniFrac) is entirely caused by an

increasing contribution of the PD component while the turnover component remains equal to 0.

Overall, the above examples (Fig. 1 and 2, Table 1) emphasize that PhyloSor$_{Turn}$ and UniFrac$_{Turn}$ are two 'narrow-sense' measures of PBD (i.e. 'true' measures of phylogenetic turnover) that are independent of total branch length difference between the two compared communities (see Fig. 2 and Table 2). Specifically, PhyloSor$_{Turn}$ and UniFrac$_{Turn}$ measure the relative magnitude of gain and loss of unique lineages between communities that is not attributable to their difference in PD (i.e. phylogenetic turnover expected if the two communities display similar levels of PD). In contrast, PhyloSor$_{PD}$ and UniFrac$_{PD}$ measure the amount of PBD caused by PD differences between phylogenetically nested communities (i.e. communities sharing at least one branch within a rooted phylogeny).

Theoretical Examples Using Simulated Communities

We simulated pairwise comparisons of communities by taking random values of *a*, *b* and *c* matching components (see formula 1 to 6) from uniform distributions between 1 and 100, where *a* is the number of species common to both communities, *b* is the number of species that occur in the first community but not in the second and *c* is the number of species that occur in the second community but not in the first. The regional species pool is thus composed of 100 species. 10 000 local communities were generated.

For each pairwise comparison, we quantified the corresponding PBD (i.e. using PhyloSor and UniFrac), and we applied the proposed decomposition of PBD. To do so, we simulated the phylogenetic relatedness among species by creating two types of regional phylogenetic trees [33], the former being generated from the PDA (proportional-to-distinguishable arrangements) model and the latter generated from the Yule model (see Fig. S1). Specifically, we aimed at testing the influence of phylogenetic tree structure (i.e. balanced vs. unbalanced trees) on the turnover and PD components of PBD. A phylogenetic tree generated under PDA (proportional-to-distinguishable arrangements) model tends to be more unbalanced than observed phylogenies because all trees with the same number of tips (i.e. species) are equally likely and the majority of potential arrangements are uneven [34,35]. Reversely, a Yule model tends to produce more balanced phylogenetic trees than empirical ones because it assumes a

Table 1. Numerical examples showing the decomposition of the total amount of phylogenetic beta diversity into two components accounting for 'true' turnover of lineages and difference in phylogenetic diversity, respectively.

	a	b	c	PhyloSor	PhyloSor$_{Turn}$	PhyloSor$_{PD}$	UniFrac	UniFrac$_{Turn}$	UniFrac$_{PD}$
A-B	5	1	1	0.166	0.166	0	0.286	0.286	0
A-C	5	1	4	0.333	0.166	0.167	0.500	0.286	0.214
A-D	5	1	5	0.375	0.166	0.209	0.545	0.286	0.259
A-E	5	1	7	0.444	0.166	0.278	0.615	0.286	0.329
A-F	5	1	8	0.474	0.166	0.308	0.643	0.286	0.357
B-C	6	0	3	0.2	0	0.2	0.333	0	0.333
B-D	6	0	4	0.25	0	0.25	0.400	0	0.400
B-E	6	0	6	0.333	0	0.333	0.500	0	0.500
B-F	6	0	7	0.368	0	0.368	0.538	0	0.538

The decomposition was based on the PhyloSor and UniFrac indices according to different pairwise comparisons involving the phylogenetic tree presented in Fig. 2. a: branch length shared by the two communities; b and c: branch length unique to the two communities compared, respectively.
doi:10.1371/journal.pone.0042760.t001

Table 2. Results showing the decomposition of PBD of coral reef fish communities into two components accounting for turnover of lineages and differences in phylogenetic diversity, respectively.

Pairwise comparisons	Decomposition based on PhyloSor						Decomposition based on UniFrac					
	PhyloSor	PhyloSor$_{Turn}$	PhyloSor$_{PD}$	SES.PhyloSor	SES.PhyloSor$_{Turn}$	SES.PhyloSor$_{PD}$	UniFrac	UniFrac$_{Turn}$	UniFrac$_{PD}$	SES.UniFrac	SES.UniFrac$_{Turn}$	SES.UniFrac$_{PD}$
Mauritius-Togian	0.44	0.32	0.12	1.07	2.96	−2.32	0.62	0.48	0.14	1.07	2.75	−2.49
Mauritius-Moorea	0.29	0.19	0.10	−0.35	0.61	−1.06	0.45	0.32	0.13	−0.33	0.64	−1.05
Mauritius-Panama	0.74	0.32	0.42	1.16	0.00	0.95	0.85	0.48	0.37	1.15	0.07	0.43
Mauritius-GBR	0.31	0.14	0.17	−1.36	−0.24	−1.03	0.47	0.25	0.22	−1.38	−0.22	−0.79
Mauritius-Vanuatu	0.33	0.26	0.07	−1.38	1.19	−2.80	0.50	0.41	0.09	−1.41	1.19	−2.61
Togian-Moorea	0.35	0.31	0.04	2.10	2.81	−1.38	0.52	0.47	0.05	2.03	2.63	−1.51
Togian-Panama	0.81	0.32	0.49	1.28	1.05	−0.62	0.90	0.49	0.41	1.27	1.07	−0.91
Togian-GBR	0.26	0.25	0.01	2.26	2.64	−0.95	0.41	0.40	0.01	2.19	2.51	−0.98
Togian-Vanuatu	0.30	0.23	0.07	2.96	1.49	1.23	0.47	0.38	0.09	2.82	1.46	0.90
Moorea-Panama	0.80	0.35	0.45	0.94	0.11	0.40	0.89	0.52	0.37	0.94	0.18	0.07
Moorea-GBR	0.23	0.17	0.06	2.12	2.39	−0.53	0.37	0.29	0.08	2.05	2.31	−0.76
Moorea-Vanuatu	0.24	0.21	0.03	1.74	2.05	−0.77	0.38	0.35	0.03	1.70	1.99	−0.86
Panama-GBR	0.82	0.35	0.47	1.14	0.42	0.03	0.91	0.52	0.39	1.14	0.49	−0.29
Panama-Vanuatu	0.79	0.35	0.44	0.40	0.33	−0.21	0.88	0.52	0.36	0.42	0.39	−0.35
GBR-Vanuatu	0.16	0.07	0.09	1.38	−1.60	3.73	0.28	0.13	0.15	1.37	−1.65	3.72

The decomposition was based on PhyloSor and UniFrac. Standardized Effect Size (SES) values were calculated from a null model (i.e. null expectation obtained by randomizing species across the tips of regional phylogenies while holding species richness and CBD constant). GBR: Great Barrier Reef (Australia).
doi:10.1371/journal.pone.0042760.t002

constant speciation/extinction rate along the tree [36,37]. Consequently, the Yule model induces a higher degree of phylogenetic similarity between species than PDA model does. The two simulated phylogenetic trees (see supplementary figures S2) were created using the R package "apTreeshape" [38].

When considering the PhyloSor index and the PDA phylogenetic tree (see Fig S2), results showed that both the turnover and PD components of PBD displayed a positive triangular relationship with PhyloSor (Fig. 3a and 3c), with an upper bound (first bisectrix) corresponding to the situation where $PhyloSor_{Turn} = PhyloSor$ (Fig. 3a) and $PhyloSor_{PD} = PhyloSor$ (Fig. 3c). $PhyloSor_{Turn} = PhyloSor$ when the two communities compared had the same phylogenetic diversity, whereas $PhyloSor_{PD} = PhyloSor$ when the two communities were completely nested in regards to their taxonomic composition (see also Fig. 2 and Table 1). When comparing $PhyloSor_{PD}$ and $PhyloSor_{Turn}$ together, we found a negative triangular relationship with an upper bound (first bisectrix) corresponding to the cases where $PhyloSor = PhyloSor_{Turn} + PhyloSor_{PD} = 1$. Similar results were obtained when considering UniFrac (Fig. 3b, 3d and 3f). Overall, these relationships based on simulated communities allow verifying the additive property of the proposed decomposition of PBD. Baselga [19] found similar triangular relationships when considering taxonomic beta diversity.

We hypothesized that phylogenetic tree topology (i.e. balance vs. unbalanced trees) may influence the observed patterns of PBD. Results showed high levels of correlation (Pearson's correlation coefficient: $r_p > 0.95$) between PBD values (PhyloSor and UniFrac) obtained using the Yule and PDA phylogenetic trees (Fig. S2). This was also verified when analysing the turnover and PD components of PBD (Fig. S2), while the levels of correlation were found to be lower (Pearson's correlation coefficient: $r_p \approx 0.8$). This
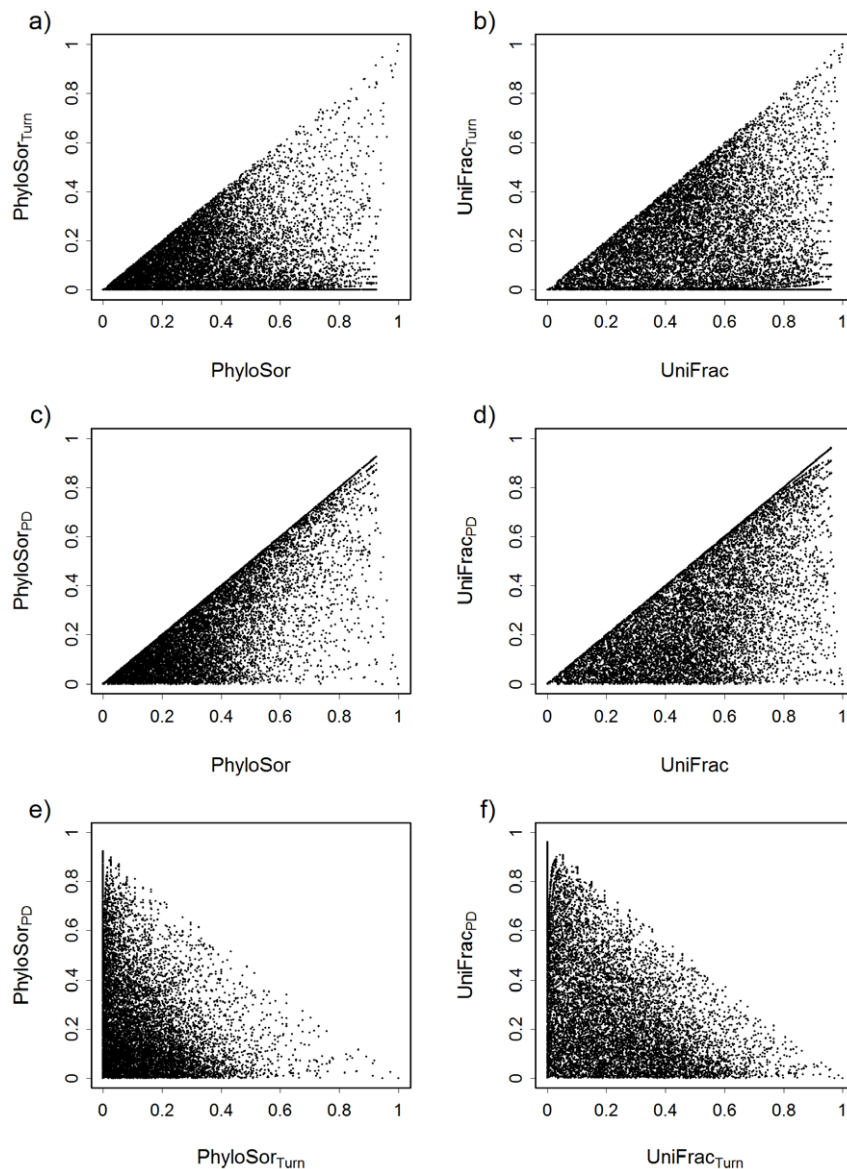

Figure 3. Relationships between phylogenetic beta diversity (PBD) and its turnover and phylogenetic diversity components. These relationships are shown according to the PhyloSor index (a, c, e) and the UniFrac index (b, d, f). Values of PBD were calculated according to simulated communities (see main text for more details).
doi:10.1371/journal.pone.0042760.g003

suggests that the shape of phylogenetic trees may have a weak influence on PBD and its turnover and PD components. However, a deeper work covering a wider panel of tree topologies [39] is needed to fully investigate the influence of phylogenetic tree shapes on PBD measurements.

Previous empirical studies emphasized that CBD and PBD may be highly correlated [3,8]. Our simulation-based approach confirmed that both the PhyloSor and UniFrac indices were highly correlated with the Sorensen (Pearson's correlation coefficient: $r_p = 0.933$, Fig. 4a) and Jaccard dissimilarity (Pearson's correlation coefficient: $r_p = 0.942$, Fig. 4b) indices, respectively. This was also verified when analysing the turnover and PD

components of PBD that showed high levels of correlation with the turnover and nestedness components of CBD (Pearson's correlation coefficient r_p ranging from 0.80 to 0.84, see Fig. 4c,d,e,f). It is worth noting that the phylogenetic diversity component of PBD is not trivially related to the nestedness component of CBD (see Fig. 4a,b). For example, when two communities are non-nested (i.e. β_{sne} or $\beta_{jne} = 0$), values of PhyloSor$_{PD}$ and UniFrac$_{PD}$ can be higher than 0. This highlights that PhyloSor$_{PD}$ and UniFrac$_{PD}$ measure the amount of PBD caused by PD differences for both nested and non-nested communities. Overall, these results emphasize that appropriate null models are required to analyze patterns of PBD and underlying processes. For instance, using
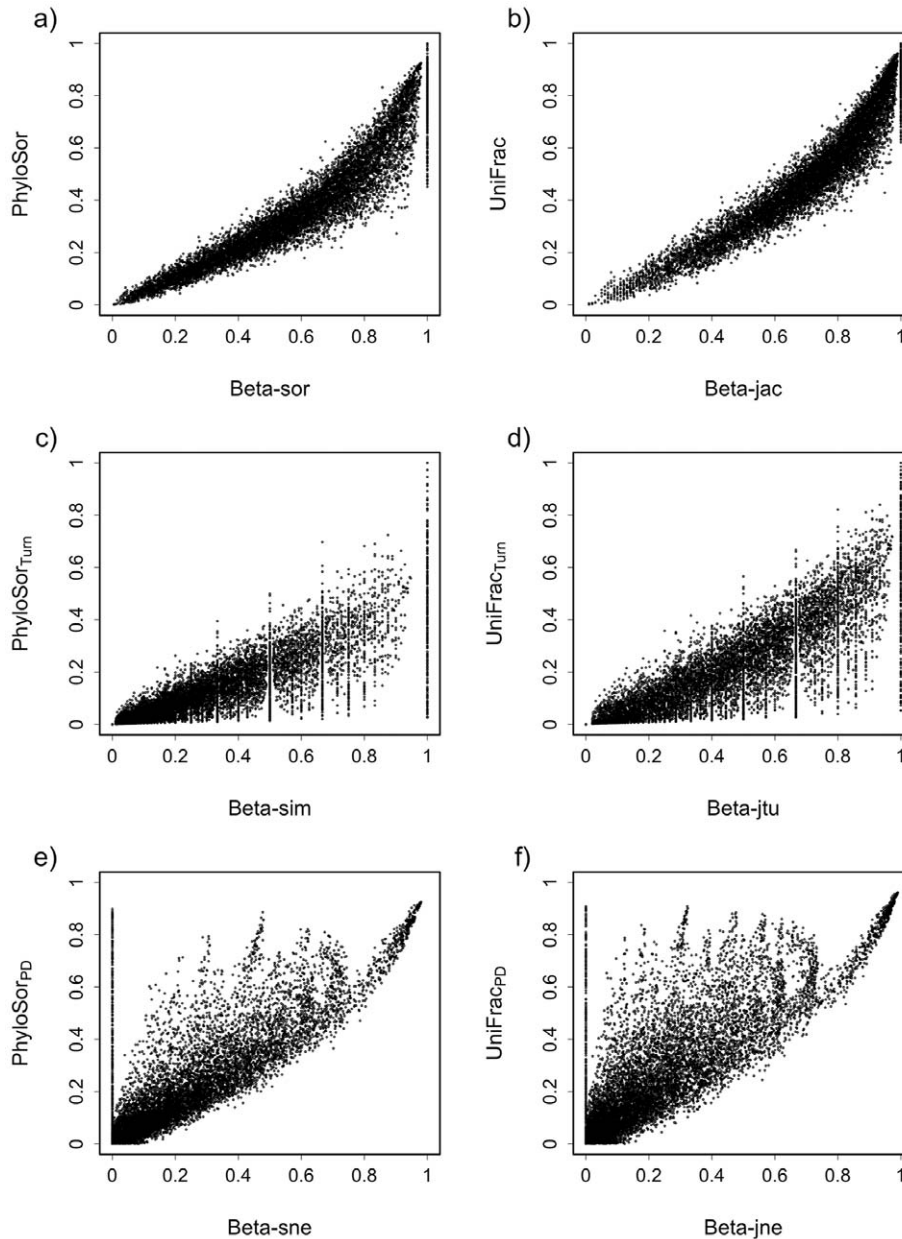

Figure 4. Relationships between phylogenetic beta diversity (PBD) and compositional beta diversity (CBD). These relationships are shown according to the Sorenson's dissimilarity index (a) and the Jaccard's dissimilarity indices (b) and their phylogenetic analogues (PhyloSor and UniFrac). Values of PBD and CBD were calculated according to simulated communities (see main text for more details). The relationship between the turnover components of PBD and CBD are also shown (c and d), as well the relationship between the phylogenetic diversity component of PBD and the nestedness component of CBD (e and f).
doi:10.1371/journal.pone.0042760.g004

PhyloSor (or Unifrac) and its turnover and PD components, one can test whether two communities are phylogenetically more or less dissimilar than what is expected given their taxa dissimilarity (CBD). This can be achieved by comparing the phylogenetic dissimilarity of the observed communities to a null expectation obtained by randomizing species across the tips of regional phylogenies while holding species richness and CBD constant [3,8].

Empirical Patterns of Phylogenetic Beta Diversity Between Coral Reef Fish Communities

We illustrated the relevance of partitioning PBD into 'true' phylogenetic turnover and PD components by exploring patterns of PBD among local communities of coral reef fishes belonging to the family Labridae. The Labridae is a species rich fish family, circa 600 species [40], that is characteristic of coral reef fish faunas around the world [41]. We compiled labrid fish species occurrences for 6 sites distributed along a longitudinal gradient (from the Indian Ocean to the Eastern Pacific passing by the Indo-Australian Archipelago, hereafter IAA, see Table S1). At each site, species occurrences were based on 12×20-min. timed swims (four locations x three habitats; the reef slope, crest and flat), to provide an overview of the local labrid fauna (census details are provided in [42]). This gradient spanned almost the entire Indian and Pacific Oceans, and encompassed the major physical factors that are thought to affect the global distribution of reef fishes [42]. To

explore PBD, we used a labrid reef fish phylogeny (108 coral reef fish species recorded from the 6 locations) that was constructed using a genetic algorithm approach based on a maximum likelihood criterion and dated using Bayesian Inference [43].

The PhyloSor index showed low levels of PBD between sites (i.e. values of PhyloSor ranging from 0.16 to 0.44, Table 2), except for the pairwise comparisons involving Panama where high levels of PBD were found (e.g. values of PhyloSor ranging from 0.74 to 0.82, see Table 2). Using UniFrac index provided similar results (Table 2). Arguably, one might conclude that high turnover of lineages occurs between Panama (East Pacific) and the other sites located in the Indian Ocean (Mauritius) and the IAA (Great Barrier Reef, Moorea, Togian and Vanuatu). However, distinguishing between the turnover and PD components of PBD showed that the level of phylogenetic turnover was roughly low for each pairwise comparison (e.g. values of $PhyloSor_{Turn}$ ranging from 0.07 to 0.35, see Table 2). In fact, the high level of PBD found between Panama and the other sites was mostly explained by their difference in PD, as shown by the $PhyloSor_{PD}$ index that ranged from 0.42 to 0.49 (Table 2, Fig. 5a). This result can be explained by the fact that Panama differs greatly from the other sites due to its low level of PD that is directly related to a low species richness (see Table S1). In contrast, the PD component of PBD showed rather low values for each between-site comparison excluding Panama (i.e. values of $PhyloSor_{PD}$ ranged from 0.01 to 0.17, Table 2, Fig. 5b), as these sites displayed comparable levels of PD (i.e. from 11.9 to 17.7, Table S1). These results suggest that
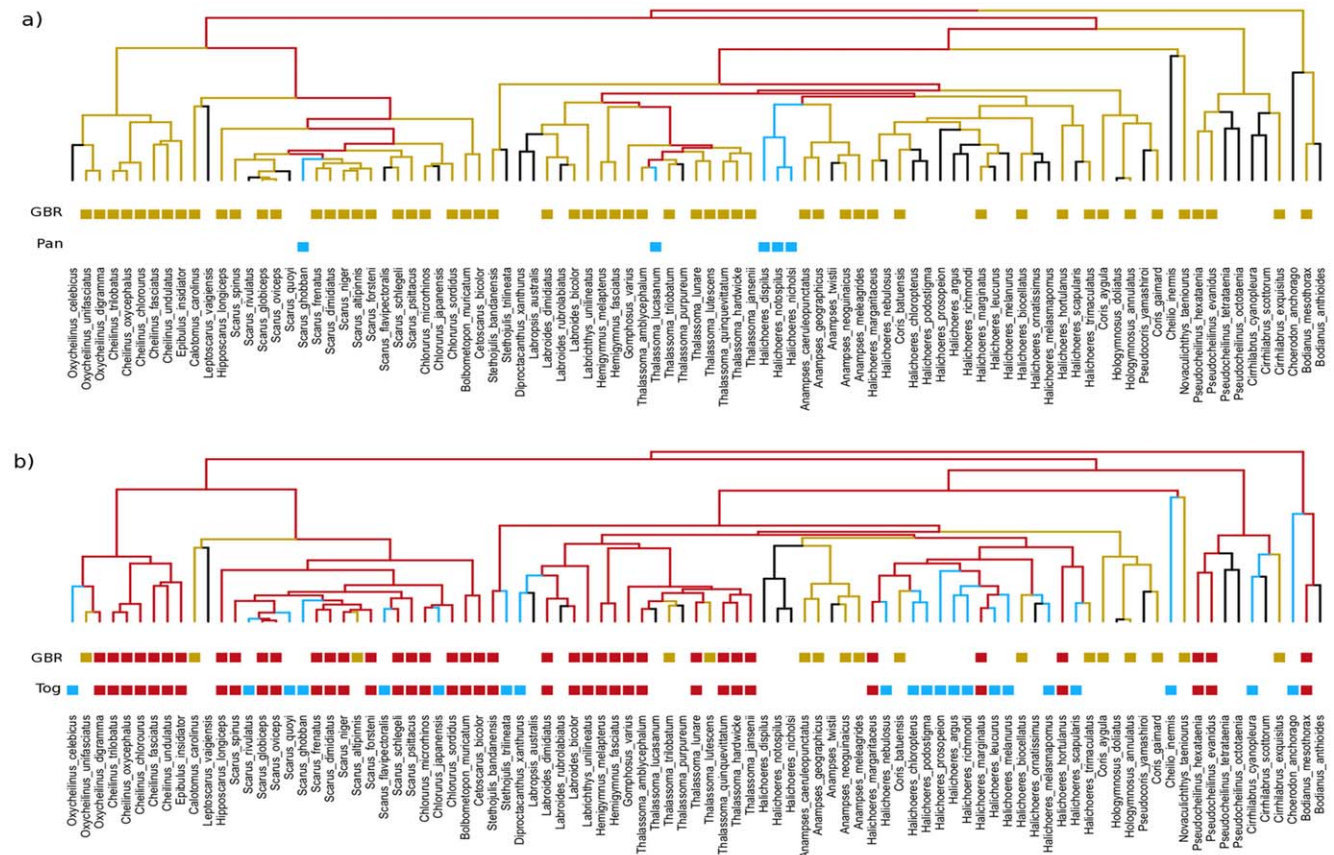

Figure 5. Phylogenetic relationships between two regional pools of labrid reef fish species. Two examples of between-region comparison are presented (a: Great Barrier Reef vs. Panama, and b: Great Barrier Reef vs. Togian). Red color shows branch length shared by the two communities compared. Green and blue colors represent the total branch length unique to each community.
doi:10.1371/journal.pone.0042760.g005

using 'broad-sense' measures of PBD such as the PhyloSor index may greatly over-estimate the level of spatial turnover of lineages if the two sites show contrasting levels of phylogenetic diversity.

Overall it appears that the IAA lies at the heart of a phylogenetic radiation, with the broader Indo-Pacific exhibiting a high degree of nestedness. In many ways this reflects the
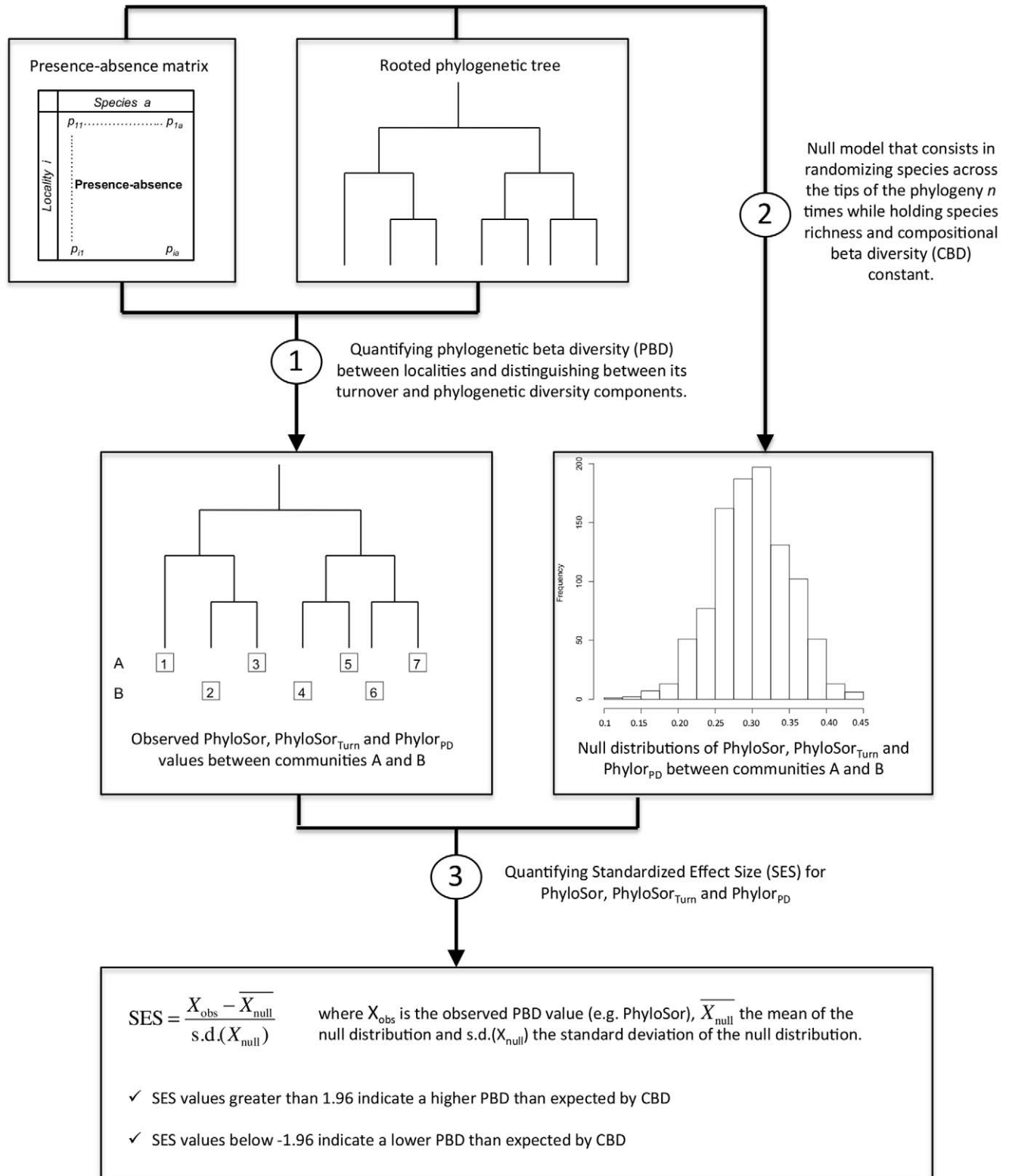

Figure 6. Flow chart of the main steps to quantify standardized effect size of phylogenetic beta diversity.
doi:10.1371/journal.pone.0042760.g006

taxonomic similarities seen in both fishes and corals across the domain [44], most of the Indo-Pacific being a low diversity subset of the IAA. A marked departure from this pattern was only found in geographically marginal locations of the East Pacific, such as Panama, where a history of isolation and species loss has resulted in an unusual low and unique diversity of coral reef fish species [44]. In addition, the development of a complex mosaic of reef habitats in the IAA during the Oligocene/Miocene has been shown to be a significant driver of cladogenesis in several coral reef fish families such as Labridae [43], hence explaining the high level of PD found in the IAA compared to Panama. Our results suggest that the high levels of PBD between Panama and the other sites is not a result of its historical association with the west Tethys [41] but with a more recent history of isolation and decline [45].

Recent approaches exploring PBD allowed the differentiation of historical (e.g. speciation and dispersal) vs. niche-based processes (e.g. environmental filtering and niche similarity) in shaping assemblage structure at both local and regional scales [3,9,14]. These studies compared observed values of PBD with those obtained by a null model where random assemblages are drawn from the overall species pool (Fig. 6). For instance, if observed values of PBD at the regional scale do not differ from what would be expected by chance alone, phylogenetic structure of regional assemblages is unlikely to be the result of historical processes [14]. In addition, as discussed above, CBD and PBD are highly correlated and an appropriate null model is therefore required to determine whether PBD is higher or lower than expected given CBD. We therefore used a null model approach (see Fig. 6) similar to that applied by Graham *et al.* [3] and Swenson *et al.* [46]. Specifically, a null distribution of PBD values was generated by randomizing species across the tips of the labrid phylogeny 9999 times while holding species richness and CBD constant. A standardized effect size (SES) was then calculated for the PhyloSor index and its turnover and PD components using the mean and standard deviation of the null distribution as follows [47]:

$$SES = \frac{X_{obs} - \overline{X_{null}}}{s.d.(X_{null})},$$

where X_{obs} is the observed PhyloSor value, $\overline{X_{null}}$ the mean of the null distribution and $s.d.(X_{null})$ the standard deviation of the null distribution. SES values greater than 1.96 indicate a higher PBD than expected by CBD while SES values below -1.96 indicate a lower PBD than expected by CBD. Specifically, we aimed at exploring whether PBD measures that account for differences in PD (PhyloSor) showed similar SES values than PBD measures that do not account for PD differences (PhyloSor$_{Turn}$).

This null model analysis showed non-random patterns of PBD for 6 pairwise comparisons out of 14 (Table 2), particularly those involving the GBR, Moorea, Togian and Vanuatu sites. However, accounting or not for PD differences between sites showed contrasting results for 3 pairwise comparisons (Mauritius-Togian, Moorea-Vanuatu, Togian-Vanuatu). For instance, the PhyloSor index showed a higher PBD than expected given CBD between the Togian and Vanuatu sites (SES = 2.96, see Table 2). This may be interpreted as a non-random spatial turnover of lineages resulting from either past speciation and extinction events or dispersal limitation of lineages or niche-based processes or a combination of both. However, controlling for PD differences between these two sites revealed a random pattern of lineage turnover (SES = 1.49). Conversely, using the PhyloSor index did not show a higher or lower PBD than expected by CBD between the Moorea and Vanuatu regions (SES = 1.04, see Table 2), while the PhyloSor$_{turn}$ index did

(SES = 2.05). Such discrepancies suggest that accounting for PD differences between sites may lead to contrasting conclusions in regards to the degree of phylogenetic structure in assemblage composition. Here our aim was not to disentangle the relative roles of ecological and evolutionary processes in shaping large-scale patterns of phylogenetic structure in coral reef fish communities. This would need analysing a larger dataset. Instead, we were interested in showing how the choice of including or not differences in PD to quantify PBD could lead to contrasting results from a null model.

Concluding Remarks

Integration of phylogenetic information into a community ecology framework has provided new insights into our understanding of the roles of ecological and evolutionary processes in shaping patterns of community structure at local and regional scales [1,2,9,48,49]. The present study participates to this emerging field of research called "ecophylogenetics" [48].

Overall, our results suggest that PD gradients may distort phylogenetic turnover patterns if the PBD measures (e.g. PhyloSor or Unifrac) incorporate PD differences between localities (or regions). This finding has important implications in the context of hypothesis testing in community ecology and biogeography [9]. For instance, one might test whether large-scale patterns of PBD can be explained by an environmental filtering process (or lineage filtering process), whereby local communities experiencing different environments contain different lineages [50]. However, these localities encompassing various regions may greatly differ in their level of phylogenetic diversity due to regional processes, e.g. regional differences in the amount of time for speciation [51] and/or differential rates of immigration [52]. Using the PhyloSor (or UniFrac) index that incorporates PD differences may hence make it difficult to distinguish between the relative roles of local-scale processes (e.g. environmental filtering) and regional processes (e.g. time for speciation) in shaping large-scale patterns of PBD. When the environmental filtering hypothesis is to be tested, we therefore recommend the use of the PhyloSor (or UniFrac) index in tandem with its 'true' phylogenetic turnover component (PhyloSor$_{turn}$), so as to control for the potential confounding effect of PD differences. Decoupling variation in beta diversity from variation in alpha diversity has rapidly emerged as an important step towards a better understanding of the drivers of community structure across latitudinal and altitudinal gradients [53,54]. From a phylogenetic perspective, our proposed decomposition of PBD into 'true' phylogenetic turnover and PD components participates to this emerging biogeographical issue.

Recent studies aimed at determining the statistical independence of several PBD metrics [11,25,55]. For instance, Swenson [11] showed that many PBD metrics (e.g. PhyloSor, UniFrac and two nearest neighbor metrics) were highly related, most of them being able to detect basal vs. terminal PBD. As mentioned by Swenson [11], future studies introducing new PBD metrics would show how these metrics actually provide novel information and strengthen the statistical toolkit of the phylogenetic community ecologist. In line with a previous additive partitioning framework of CBD [19,26], we provided new insights into a specific class of PBD metrics that belongs to the family of phylogenetic diversity-based dissimilarity measures. Specifically, we propose a new phylogenetic turnover metric that is independent of variation in PD between localities (or regions). We hope that our proposed PBD metrics will help future studies to unravel the mechanisms driving large-scale patterns of biodiversity.

Supporting Information

Figure S1 Simulated phylogenetic trees obtained by the Yule (a) and PDA (b) models.
(TIFF)

Figure S2 Relationships between the PDB values obtained using two different phylogenetic trees (Yule vs. PDA model).
(TIF)

Table S1 Values of phylogenetic diversity (calculated as the total sum of branch length) and species richness of coral reef fish species belonging to the family of Labridae for each studied site.
(DOCX)

File S1 R code for quantifying phylogenetic beta diversity and its 'true' turnover and phylogenetic diversity components (requires R; download R software from http://cran.r-project.org).
(R)

File S2 R code to run the example illustrated in Table 1 and Figure 2 (requires R; download R software from http://cran.r-project.org).
(R)

File S3 Species occurrence matrix used in Table 1 and Figure 2.
(CSV)

File S4 Hypothetical phylogenetic tree used in Table 1 and Figure 2.
(NWK)

References

1. Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. Annu Rev Ecol Syst 33: 475–505.
2. Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. Ecol Lett 12: 693–715.
3. Graham CH, Parra JL, Rahbek C, McGuire JA (2009) Phylogenetic structure in tropical hummingbird communities. Proc Natl Acad Sci U S A 106: 19673–19678.
4. Cadotte MW, Cardinale BJ, Oakley TH (2008) Evolutionary history and the effect of biodiversity on plant productivity. Proc Natl Acad Sci U S A 105: 17012–17017.
5. Faith DP, Magallon S, Hendry AP, Conti E, Yahara T, et al. (2010) Evosystem services: an evolutionary perspective on the links between biodiversity and human well-being. Current Opinion in Environmental Sustainability 2: 66–74.
6. Cadotte MW, Davies TJ (2010) Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. Divers Distrib 16: 376–385.
7. Mouillot D, Albouy C, Guilhaumon F, Lasram FBR, Coll M, et al. (2011) Protected and Threatened Components of Fish Biodiversity in the Mediterranean Sea. Curr Biol 21: 1044–1050.
8. Bryant JA, Lamanna C, Morlon H, Kerkhoff AJ, Enquist BJ, et al. (2008) Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. Proc Natl Acad Sci U S A 105: 11505–11511.
9. Graham CH, Fine PVA (2008) Phylogenetic diversity: linking ecological and evolutionary processes across space in time. Ecol Lett 11: 1265–1277.
10. Morlon H, Schwilk DW, Bryant JA, Marquet PA, Rebelo AG, et al. (2011) Spatial patterns of phylogenetic diversity. Ecol Lett 14: 141–149.
11. Swenson NG (2011) Phylogenetic Beta Diversity Metrics, Trait Evolution and Inferring the Functional Beta Diversity of Communities. PLoS One 6: e21264.
12. Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71: 8228–8235.
13. Faith DP, Lozupone CA, Nipperess D, Knight R (2009) The Cladistic Basis for the Phylogenetic Diversity (PD) Measure Links Evolutionary Features to Environmental Gradients and Supports Broad Applications of Microbial Ecology's "Phylogenetic Beta Diversity" Framework. Int J Mol Sci 10: 4723–4741.
14. Gomez JP, Bravo GA, Brumfield RT, Tello JG, Cadena CD (2010) A phylogenetic approach to disentangling the role of competition and habitat filtering in community assembly of Neotropical forest birds. J Anim Ecol 79: 1181–1192.
15. Nipperess DA, Faith DP, Barton K (2010) Resemblance in phylogenetic diversity among ecological assemblages. J Veg Sci 21: 809–820.
16. Koleff P, Gaston KJ, Lennon JJ (2003) Measuring beta diversity for presence-absence data. J Anim Ecol 72: 367–382.
17. Baselga A (2007) Disentangling distance decay of similarity from richness gradients: response to Soininen et al. 2007. Ecography 30: 838–841.
18. Leprieur F, Olden JD, Lek S, Brosse S (2009) Contrasting patterns and mechanisms of spatial turnover for native and exotic freshwater fish in Europe. J Biogeogr 36: 1899–1912.
19. Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity. Glob Ecol Biogeogr 19: 134–143.
20. Svenning JC, Flojgaard C, Baselga A (2011) Climate, history and neutrality as drivers of mammal beta diversity in Europe: insights from multiscale deconstruction. J Anim Ecol 80: 393–402.
21. Leprieur F, Tedesco PA, Hugueny B, Beauchard O, Durr HH, et al. (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. Ecol Lett 14: 325–334.
22. Hortal J, Diniz JAF, Bini LM, Rodriguez MA, Baselga A, et al. (2011) Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. Ecol Lett 14: 741–748.
23. Andrew ME, Wulder MA, Coops NC, Baillargeon G (2012) Beta-diversity gradients of butterflies along productivity axes. Glob Ecol Biogeogr 21: 352–364.
24. Baselga A, Gómez-Rodríguez C, JM L (2012) Historical legacies in world amphibian diversity revealed by the turnover and nestedness components of beta diversity. PLoS One 7: e32341.
25. Ives AR, Helmus MR (2010) Phylogenetic Metrics of Community Similarity. Am Nat 176: E128–E142.
26. Baselga A (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. Glob Ecol Biogeogr In press.
27. Gaston KJ, Davies RG, Orme CDL, Olson VA, Thomas GH, et al. (2007) Spatial turnover in the global avifauna. Proc Biol Sci 274: 1567–1574.
28. Lennon JJ, Koleff P, Greenwood JJD, Gaston KJ (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. J Anim Ecol 70: 966–979.
29. McKnight MW, White PS, McDonald RI, Lamoreux JF, Sechrest W, et al. (2007) Putting beta-diversity on the map: Broad-scale congruence and coincidence in the extremes. PLoS Biology 5: 2424–2432.
30. Rodrigues ASL, Gaston KJ (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas. Biol Conserv 105: 103–111.
31. Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61: 1–10.
32. Team RDC (2011) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
33. Mouchet MA, Mouillot D (2011) Decomposing phylogenetic entropy into alpha, beta and gamma components. Biol Lett 7: 205–209.
34. Rosen DE (1978) Vicariant patterns and historical explanation in biogeography. Syst Zool 27: 159–188.
35. Cunningham SA (1995) Problems with null models in the study of phylogenetic radiation. Evolution 49: 1292–1294.
36. Yule GU (1925) A mathemahcal theory of evolution, based on the conclusions of Dr J C Willis, F R S. Philos Trans R Soc Lond B Biol Sci 213: 21–87.
37. Mooers AO, Heard SB (1997) Evolutionary process from phylogenetic tree shape. Q Rev Biol 72: 31–54.
38. Bortolussi N, Durand E, Blum M, Francois O (2006) apTreeshape: statistical analysis of phylogenetic tree shape. Bioinformatics 22: 363–364.
39. Schweiger O, Klotz S, Durka W, Kuhn I (2008) A comparative test of phylogenetic diversity indices. Oecologia 157: 485–495.
40. Parenti P, Randall JE (2011) Checklist of the species of the families Labridae and Scaridae : an update. Smithiana Bulletin 13: 29–44.
41. Bellwood DR, Wainwright PC (2002) The history and biogeography of fishes on coral reels. In: Sale PF, editor. Coral Reef Fishes: dynamic and diversity in a complex ecosystem. San Diego, CA, , USA: Elsevier. pp. 5–32.
42. Bellwood DR, Hughes TP, Connolly SR, Tanner J (2005) Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. Ecol Lett 8: 643–651.
43. Cowman PF, Bellwood DR (2011) Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. J Evol Biol 24: 2543–2562.

44. Bellwood DR, Hughes TP (2001) Regional-scale assembly rules and biodiversity of coral reefs. Science 292: 1532–1534.
45. O'Dea A, Jackson JBC, Fortunato H, Smith JT, D'Croz L, et al. (2007) Environmental change preceded Caribbean extinction by 2 million years. Proc Natl Acad Sci U S A 104: 5501–5506.
46. Swenson NG, Anglada-Cordero P, Barone JA (2011) Deterministic tropical tree community turnover: evidence from patterns of functional beta diversity along an elevational gradient. Proc Biol Sci 278: 877–884.
47. Gotelli NJ, Graves GR (1996) Null models in ecology. Washington, D.C.: Smithsonian Institution Press. 368 p.
48. Mouquet N, Devictor V, Meynard CN, Munoz F, Bersier LF, et al. (2012) Ecophylogenetics: advances and perspectives. Biol Rev In press.
49. Swenson NG, Erickson DL, Mi X, Bourg NA, Forero-Montana J, et al. (2012) Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. Ecology In press.
50. Pommier T, Douzery EJP, Mouillot D (2012) Environment drives high phylogenetic turnover among oceanic bacterial communities. Biol Lett In press.
51. Stevens RD (2011) Relative effects of time for speciation and tropical niche conservatism on the latitudinal diversity gradient of phyllostomid bats. Proc Biol Sci 278: 2528–2536.
52. Davies TJ, Buckley LB (2011) Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. Philos Trans R Soc Lond B Biol Sci 366: 2414–2425.
53. Kraft NJB, Comita LS, Chase JM, Sanders NJ, Swenson NG, et al. (2011) Disentangling the Drivers of beta Diversity Along Latitudinal and Elevational Gradients. Science 333: 1755–1758.
54. Chase JM, Kraft NJB, Smith KG, Vellend M, Inouye BD (2011) Using null models to disentangle variation in community dissimilarity from variation in alpha-diversity. Ecosphere 2: Article 24.
55. Feng G, Zhang JL, Pei NC, Rao MD, Mi XC, et al. (2012) Comparison of phylobetadiversity indices based on community data from Gutianshan forest plot. Chin Sci Bull 57: 623–630.