SDA*: A Simple and Unifying Solution to Recent Bioinformatic Challenges for Conservation Genetics

Bui Quang Minh
Center for Integrative Bioinformatics
Max F. Perutz Laboratories
University of Vienna
Medical University of Vienna
Veterinary University of Vienna
Vienna, Austria
minh.bui@mfpl.ac.at

Steffen Klaere
Department of Mathematics
University of Auckland
Auckland, New Zealand
steffen.klaere@gmail.com

Arndt von Haeseler
Center for Integrative Bioinformatics
Max F. Perutz Laboratories
University of Vienna
Medical University of Vienna
Veterinary University of Vienna
Vienna, Austria
arndt.von.haeseler@mfpl.ac.at

Abstract—Recently, several algorithms have been proposed to tackle different conservation questions under phylogenetic diversity. Such questions are variants of the more general problem of budgeted reserve selection under split diversity, an NP-hard problem. Here, we present a novel framework, Split Diversity Algorithm* (SDA*), to unify all these attempts. More specifically, SDA* transforms the budgeted reserve selection problem into a binary linear programming (BLP), that can be solved by available linear optimization techniques. SDA* guarantees to find optimal solutions in reasonable time.

Keywords-conservation genetics; phylogenetic diversity; split diversity; linear programming; phylogenetic network; split system

I. INTRODUCTION

The importance of phylogenetic information in bio-diversity assessment and conservation biology has been intensively discussed since the early 1990s [1]–[4]. Vane-Wright et al. [5] were the first to include the taxonomy of species into biodiversity evaluation and introduced the concept of *taxic diversity*. Faith [6] and Crozier [7] soon recognized that taxic diversity can be improved by taking into account evolutionary time. Consequently, Faith defined the *phylogenetic diversity (PD)* of a set of taxa as the sum of the edge lengths of the minimal sub-tree connecting these taxa with the root. PD has subsequently been extended to account for species extinction risks [8], [9], budget constraints [10], and ecological interactions [11]. Recently, it has been observed that the traditional measure of *species richness* [12] is a good surrogate for PD [13], [14]. However, a recent study on the distribution of PD for the plants of the Cape of South Africa showed that PD is decoupled from species richness [15].

With the advent of molecular data, phylogenetic trees are nowadays often inferred from several genomic regions [16]. Such phylogenomic methods, however, pose a problem for PD [17], [18] because it was well documented that different genes exhibit different rates of evolution [19] or conflicting phylogenetic signals [20]. Even single gene trees may be unreliable due to model misspecification or violations of treelikeness in the data [21]. These may cause unreliable PD estimates or different PD values computed from different trees. To resolve this issue we have recently introduced *split diversity (SD)* [22]. SD combines PD from different trees, thus integrating incompatible phylogenetic information into the analysis. It has been shown that SD can also be computed on the *split system* [23] representing the given collection of trees [22], [24]. SD equals PD when the split system corresponds to a tree, thus generalizes PD.

Recently, various conservation questions using PD have gained increasing interest, particularly with respect to their associated bioinformatic challenges. The simplest question is *taxon selection under PD*: For a given set X of taxa, its phylogenetic tree and a number k find k taxa which maximize the PD over all possible k-element subsets [6]. This question is motivated by the fact that limited resources can only support a fraction of all species, thus an optimal k-set will maximize the protection of PD and feature diversity. It was shown that the simple greedy strategy guarantees the optimal solution [25], [26] and efficient algorithms exist [17], [24]. This simple question has been extended to address more realistic scenarios that we discuss below.

First, taxon selection under PD naturally extends to SD and becomes *taxon selection under SD*: For a given set X of taxa, its split system (or collection of trees), and a number k, find a subset of k taxa which maximizes the SD over all taxon subsets of size k. Under SD, the greedy algorithm no longer works [17], [18]. For general split systems taxon selection under SD was proven NP-hard [24]. However, for special cases such as circular and affine split systems [22], [24] and split systems of two trees [27], efficient algorithms exist.

The second level of complexity is *budgeted taxon selection* (under either PD or SD): Preserving each taxon comes at a cost and given a limited total budget, select those taxa that maximize PD/SD within the allotted budget (budget constraints). Taxon selection is a special case of budgeted taxon selection assuming equal conservation costs among taxa [28]. Efficient dynamic programming algorithms were presented for budgeted taxon selection under PD [28], [29] and under SD for circular split systems [29]. However, no algorithm is known for general split systems.

The third scenario to include more realistic aspects

33

is *reserve selection*: Assuming that the habitat of the taxa is partitioned into several geographical areas, find k areas such that the PD/SD of the taxa present in the k areas is maximized. The selected areas might help in the establishment of nature reserves. Mathematically, taxon selection is a special case of reserve selection by introducing artificial areas containing single taxa. Reserve selection is also NP-hard [18].

With the aforementioned extensions one may ask if it is possible to solve the generalized problem that considers all extensions at the same time, thus providing a unifying framework for any possible conservation question. This most general problem is called *budgeted reserve selection under SD*: Given a set of areas, each with a conservation cost, and a total budget find a subset of areas with maximal SD whose sum of costs do no exceed the total budget. This general problem is of course NP-hard. An approximate algorithm [30] was given for budgeted reserve selection under PD. Here, we present a simple and unifying framework (SDA*) to tackle budgeted reserve selection under SD that works for arbitrary split systems. To this end, we transform budgeted reserve selection under SD into a *binary linear programming (BLP)* problem [31], which is then solved with available software packages.

The paper is organized as follows. First, we briefly describe the measure of split diversity. Second, we explain how to transform the budgeted reserve selection under SD into a BLP. We finally discuss the possibility to include more extended conservation scenarios into the SDA* approach.

II. THE INDEX OF SPLIT DIVERSITY

Here we briefly introduce *split diversity* [22]. Let X be a set of taxa. A *split* σ on X is defined as a bipartition of X into two disjoint and complementary subsets. Splits are normally denoted by $\sigma = A|B$ where $B \equiv X \setminus A$. A *split system* Σ is simply a collection of splits on X. For a *weighted split system* (Σ, λ) the weight function λ assigns each $\sigma \in \Sigma$ a non-negative weight $\lambda(\sigma)$. Fig. 1 gives an example of a weighted split system.

We define the split diversity of a taxon subset $S \subset X$, $\mathrm{sd}(S)$, on a weighted split system (Σ, λ) as

$$\mathrm{sd}(S) = \sum_{\substack{\sigma = A|B \in \Sigma: \\ A \cap S \neq \emptyset, B \cap S \neq \emptyset}} \lambda(\sigma). \tag{1}$$

That means the sum is over all splits of Σ that separate at least two taxa of S.

III. SDA*: A BINARY LINEAR PROGRAMMING APPROACH FOR BUDGETED RESERVE SELECTION UNDER SD

Linear programming (LP) and particularly binary linear programming (BLP) are powerful mathematical tools with a wide range of practical applications [31], [32]. Here we will utilize BLP to solve the budgeted reserve selection under SD.

Assuming a taxon set X together with a weighted split system (Σ, λ) on X. Moreover, suppose that the natural

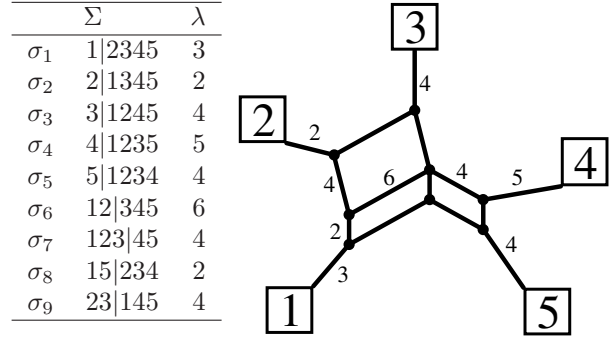|	Σ	λ
σ_1	1|2345	3
σ_2	2|1345	2
σ_3	3|1245	4
σ_4	4|1235	5
σ_5	5|1234	4
σ_6	12|345	6
σ_7	123|45	4
σ_8	15|234	2
σ_9	23|145	4

Figure 1. An example weighted split system (Σ, λ) for five taxa and its graphical representation, split network, redrawn from [22]. Weighted split systems generalize phylogenetic trees by allowing for incompatible phylogenetic signals (See [22] for more details).

habitat of the taxa in X is partitioned into m (geographical) areas $R_1, R_2, \ldots, R_m$. We denote the collection of these areas as $\mathcal{R} = \{R_1, R_2, \ldots, R_m\}$. In the following, it suffices to represent each area with the taxa in X that occur in the area. Hence, we will not distinguish between area R_i and its taxon set, that is $R_i \subset X$ for all $i = 1, \ldots, m$. For an area-subset $\mathcal{W} \subset \mathcal{R}$ we define the split diversity of $\mathcal{W}$ as the SD score of the taxa living in at least one area of $\mathcal{W}$:

$$\mathrm{sd}(\mathcal{W}) = \mathrm{sd}\left(\bigcup_{R \in \mathcal{W}} R\right) = \sum_{\sigma} \lambda(\sigma), \tag{2}$$

where the summation is over all splits $\sigma = A|B \in \Sigma$ such that at least one taxon from A and another taxon from B must be present in $\bigcup_{R \in \mathcal{W}} R$. Such splits σ are called to be *preserved* in $\mathcal{W}$.

The budgeted reserve selection under SD problem is then: Given a set X of taxa, a weighted split system (Σ, λ), an area collection $\mathcal{R} = \{R_1, \ldots, R_m\}$ with associated conservation costs $c_1, \ldots, c_m$, and a total budget B, select an area-subset $\mathcal{W}_{\max}$ with maximal SD subject to:

$$\sum_{i:R_i \in \mathcal{W}_{\max}} c_i \leq B \tag{3}$$

A. Transformation into a BLP

Obviously the evaluation of all 2^m area combinations to solve the problem is computationally not feasible for large m. In the following we will explain how to transform budgeted reserve selection under SD into a BLP problem. We illustrate this with the example weighted split system (Σ, λ) comprising nine splits $\sigma_1, \ldots, \sigma_9$ (Fig. 1) and four hypothetical areas $R_1 = \{1, 5\}, R_2 = \{2, 4\}, R_3 = \{3, 5\}, R_4 = \{1, 2, 3\}$.

The central idea is to encode any area-subset $\mathcal{W}$ as an m-element binary vector $(x_1, x_2, \ldots, x_m)$. The x_i $(1 \leq i \leq m)$ are set to 1 if $R_i \in \mathcal{W}$ and 0 otherwise. Therefore, x_i acts as an indicator variable for the ith area in $\mathcal{W}$. We call x_i area-variables. Moreover, for each split σ_j we introduce a split-variable y_j: $y_j = 1$ if σ_j is preserved in $\mathcal{W}$, and $y_j = 0$ otherwise. Therefore, our example has nine split-variables $y_1, y_2, \ldots, y_9$. With the split-variables

Table I
INTERSECTION OF FOUR AREAS WITH TWO SIDES OF SPLIT σ_8. A CROSS INDICATES NON-EMPTY INTERSECTION.

$\cap$	σ_8	
	$\{1,5\}$	$\{2,3,4\}$
$R_1 = \{1,5\}$	$\times$	
$R_2 = \{2,4\}$		$\times$
$R_3 = \{3,5\}$	$\times$	$\times$
$R_4 = \{1,2,3\}$	$\times$	$\times$

we can rewrite (2) as:

$$\mathrm{sd}(\mathcal{W}) = \sum_{j=1}^{9} \lambda(\sigma_j) y_j. \qquad (4)$$

This is the objective function to be maximized.

We need now a way to determine the preserved splits for arbitrary $\mathcal{W}$. With respect to the introduced variables this reduces to determining the values of split-variables y_j based on the area-variables x_i. We will show that the preservedness can be computed independently of the generation of all possible $\mathcal{W}$ by comparing each split to the taxon composition of each area. For example, split $\sigma_8 = 15|234$ is preserved in $\mathcal{W}$ (or $y_8 = 1$) if and only if at least one taxon from $\{1,5\}$ and one taxon from $\{2,3,4\}$ are present in $\mathcal{W}$. Table I shows that this is the case, if at least one area from $\{R_1, R_3, R_4\}$ and another area of $\{R_2, R_3, R_4\}$ are included in $\mathcal{W}$. With regards to the introduced variables, $y_8 = 1$ if and only if $x_1 + x_3 + x_4 \geq 1$ and $x_2 + x_3 + x_4 \geq 1$. On the other hand, we can express these conditions by two inequalities:

$$y_8 \leq x_1 + x_3 + x_4, \\ y_8 \leq x_2 + x_3 + x_4. \qquad (5)$$

The reason is if the right-hand side of any inequalities in (5) is zero, y_8 will be zero. If both sums are ≥ 1, maximizing the objective function (4) will effectively return $y_8 = 1$. Thus, we have established the equivalence of (5) with the boundary condition imposed on y_8. We call (5) the split constraint inequalities for y_8. This can be similarly done for every split.

Finally, it is easy to see that the budget constraint (3) is equivalent to:

$$\sum_{i=1}^{m} c_i x_i \leq B. \qquad (6)$$

Taking the above arguments together, budgeted reserve selection under SD becomes maximizing function (4) on binary variables given the transformed budget constraint (6) and a series of split constraints for $y_1, \ldots, y_9$ (Table II). This is clearly a BLP problem. To solve the transformed BLP problem one can apply exact branch-and-bound algorithms [32]. Theoretically, the algorithms have exponential time complexity. However, their implementations (e.g., the CPLEX package or Gurobi optimizer) work very well for large BLP systems.

B. Accelerating SDA*

We denote that the performance of the BLP solver depends on the number of (binary) variables and the

Table II
THE TRANSFORMED BLP PROBLEM OF THE BUDGETED RESERVE SELECTION UNDER SD.

maximize:	
$3y_1 + 2y_2 + 4y_3 + 5y_4 + 4y_5 + 6y_6 + 4y_7 + 2y_8 + 4y_9$	
subject to:	
$c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 \leq B$	(budget constraint)
$y_1 \leq x_1 + x_4$	
$y_1 \leq x_1 + x_2 + x_3 + x_4$	
$y_2 \leq x_2 + x_4$	
$y_2 \leq x_1 + x_2 + x_3 + x_4$	
$y_3 \leq x_3 + x_4$	
$y_3 \leq x_1 + x_2 + x_3 + x_4$	
$y_4 \leq x_2$	
$y_4 \leq x_1 + x_2 + x_3 + x_4$	
$y_5 \leq x_1 + x_3$	
$y_5 \leq x_1 + x_2 + x_3 + x_4$	(split constraints)
$y_6 \leq x_1 + x_2 + x_4$	
$y_6 \leq x_1 + x_2 + x_3 + x_4$	
$y_7 \leq x_1 + x_2 + x_3$	
$y_7 \leq x_1 + x_2 + x_3 + x_4$	
$y_8 \leq x_1 + x_3 + x_4$	
$y_8 \leq x_2 + x_3 + x_4$	
$y_9 \leq x_2 + x_3 + x_4$	
$y_9 \leq x_1 + x_2 + x_3 + x_4$	
$x_1, \ldots, x_4 \in \{0,1\}$	(binary constraints)
$y_1, \ldots, y_9 \in \{0,1\}$	

number of split constraints. First, we observe that due to the special structure of the split constraints (Table II), only area-variables x_i need to be binary. Since all inequalities for split constraints have a coefficient of 1 and the objective function solely acts on split variables y_j, all returned solutions for y_j are also binary. Consequently, we have a so-called mixed integer linear programming problem (MILP) [32] where x_i are binary variables and y_j are real variables of the range $[0,1]$. This has a great effect in practice because the computational time of the MILP solver depends heavily on the number of binary variables, thus neglecting the influence of the (possibly large) number of y_j.

Second, those split constraints of the form $y_j \leq x_1 + x_2 + x_3 + x_4$ (8 constraints in Table II) are always satisfied if the total budget B is enough to cover at least one area, i.e., $B \geq \min c_i$. Hence, such constraints are redundant and can be dropped from the system.

Third, a common technique is to solve the MILP problem in two phases. In the first phase, we solve the "non-integer" LP system by allowing all x_i to have real values in the range $[0,1]$. This is done very efficiently with the simplex or the interior point method [32]. In the second stage, we check whether the returned solution contain only binary values for every x_i or not. If the former is true, we are done. Otherwise, we have to re-solve the MILP system.

We have implemented SDA* in the PDA software (http://www.cibiv.at/software/pda) employing the Gurobi optimizer. We tested the performance on a split system of 700 taxa and 6000 splits with 200 areas and under various budget constraints. The results showed an average running time of 1 minute per transformed MILP problem on a CPU of 2.66 GHz. SDA* is therefore efficient for large data sets.

IV. Discussions

BLP has previously been applied for reserve selection under taxon richness and PD [33]–[35]. Since SD is a generalization of PD and taxon richness, the proposed SDA* framework generalizes all these LP-based approaches. Despite its simplicity, SDA* provides a unifying solution to all recent attempts dealing with subtypes of the most general problem of budgeted reserve selection under SD.

Moreover, SDA* allows us to summarize PD from trees being sampled from a bootstrap analysis, thus accounting for uncertainties on the data or tree reconstruction methods. Similarly SDA* can also be used with the trees produced by Bayesian analyses. Crozier et al. [36], [37] also pursued a bootstrap approach, but they were interested in a different summary statistic, i.e., confidence interval estimates of PD for single regions.

Greedy algorithms for reserve selection have frequently been used under the name *complementarity principle* [5], [38]. However, it has been criticized for sub-optimality [33], i.e., the areas obtained may not contain the maximal diversity. SDA* guarantees to provide the optimal selection in reasonable time.

Recently, LP has been employed to optimizing PD under ecological interactions such as prey-predator relationships [39]. We suspect that such an approach can also be applied for SD. Other constraints such as requiring that each taxon is present in at least two selected areas can also be expressed as linear inequalities. SDA* will therefore prove to be a versatile bioinformatic tool for solving conservation questions.

Acknowledgment

References

[1] R. M. May, "Taxonomy as destiny," *Nature*, vol. 347, pp. 129–130, 1990.

[2] C. J. Humphries, P. H. Williams, and R. I. Vane-Wright, "Measuring Biodiversity Value for Conservation," *Annu. Rev. Ecol. Syst.*, vol. 26, pp. 93–111, 1995.

[3] R. H. Crozier, "Preserving the information content of species: Genetic diversity, phylogeny, and conservation worth," *Annu. Rev. Ecol. Syst.*, vol. 28, pp. 243–68, 1997.

[4] J. C. Avise, "Colloquium paper: Three ambitious (and rather unorthodox) assignments for the field of biodiversity genetics," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. Supplement_1, pp. 11 564–11 570, 2008.

[5] R. I. Vane-Wright, C. J. Humphries, and P. H. Williams, "What to protect? - systematics and the agony of choice," *Biol. Conserv.*, vol. 55, pp. 235–254, 1991.

[6] D. P. Faith, "Conservation Evaluation and Phylogenetic Diversity," *Biol. Conserv.*, vol. 61, pp. 1–10, 1992.

[7] R. H. Crozier, "Genetic diversity and the agony of choice," *Biol. Conserv.*, vol. 61, pp. 11–15, 1992.

[8] M. L. Weitzman, "On diversity," *Q. J. Econ.*, vol. 107, pp. 363–405, 1992.

[9] L. Witting and V. Loeschcke, "The optimization of biodiversity conservation," *Biol. Conserv.*, vol. 71, pp. 205–207, 1995.

[10] M. L. Weitzman, "The Noah's Ark problem," *Econometrica*, vol. 66, no. 6, pp. 1279–1298, 1998.

[11] L. Witting, J. Tomiuk, and V. Loeschcke, "Modelling the optimal conservation of interacting species," *Ecol. Model.*, vol. 125, no. 2-3, pp. 123–144, 2000.

[12] K. J. Gaston and J. I. Spicer, *Biodiversity: An Introduction*, 2nd ed. Blackwell Publishing Professional, 2004.

[13] S. Polasky, B. Csuti, C. A. Vossler, , and S. M. Meyers, "A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for north american birds," *Biol. Conserv.*, vol. 97, no. 1, pp. 99–105, 2001.

[14] A. S. L. Rodrigues, T. M. Brooks, and K. J. Gaston, "Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference?" in *Phylogeny and Conservation*, A. Purvis, J. L. Gittleman, and T. Brooks, Eds. Cambridge, UK: Cambridge University Press, 2005, pp. 101–119.

[15] F. Forest, R. Grenyer, M. Rouget, T. J. Davies, R. M. Cowling, D. P. Faith, A. Balmford, J. C. Manning, S. Proches, M. van der Bank, G. Reeves, T. A. J. Hedderson, and V. Savolainen, "Preserving the evolutionary potential of floras in biodiversity hotspots," *Nature*, vol. 445, pp. 757–760, 2007.

[16] F. Delsuc, H. Brinkmann, and H. Philippe, "Phylogenomics and the reconstruction of the tree of life," *Nat. Rev. Genet.*, vol. 6, no. 5, pp. 361–375, 2005.

[17] B. Q. Minh, S. Klaere, and A. von Haeseler, "Phylogenetic diversity within seconds," *Syst. Biol.*, vol. 55, no. 5, pp. 769–773, 2006.

[18] V. Moulton, C. Semple, and M. Steel, "Optimizing phylogenetic diversity under constraints," *J. Theor. Biol.*, vol. 246, pp. 186–194, 2007.

[19] D. Graur and W.-H. Li, *Fundamentals of Molecular Evolution*, 2nd ed. Sunderland, Massachusetts: Sinauer Associates, 2000.

[20] M. Nei, *Molecular Evolutionary Genetics*. New York: Columbia University Press, 1987.

[21] L. S. Jermiin, V. Jayaswal, F. Ababneh, and J. Robinson, "Phylogenetic model evaluation," in *Bioinformatics - Volume I: Data, Sequences Analysis and Evolution*, J. Keith, Ed. Totowa, NJ: Humana Press, 2008, pp. 331–363.

[22] B. Q. Minh, S. Klaere, and A. von Haeseler, "Taxon selection under split diversity," *Syst. Biol.*, vol. 58, no. 6, pp. 586–594, 2009.

[23] H.-J. Bandelt and A. W. M. Dress, "Split decomposition: A new and useful approach to phylogenetic analysis of

distance data," *Mol. Phylogenet. Evol.*, vol. 1, pp. 242–252, 1992.

[24] A. Spillner, B. T. Nguyen, and V. Moulton, "Computing phylogenetic diversity for split systems," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 5, no. 2, pp. 235–244, 2008.

[25] M. Steel, "Phylogenetic Diversity and the Greedy Algorithm," *Syst. Biol.*, vol. 54, no. 4, pp. 527–529, 2005.

[26] F. Pardi and N. Goldman, "Species choice for comparative genomics: Being greedy works," *PLoS Genet.*, vol. 1, pp. 672–675, 2005.

[27] M. Bordewich, C. Semple, and A. Spillner, "Optimizing phylogenetic diversity across two trees," *Appl. Math. Lett.*, vol. 22, no. 5, pp. 638–641, 2009.

[28] F. Pardi and N. Goldman, "Resource aware taxon selection for maximizing phylogenetic diversity," *Syst. Biol.*, vol. 56, no. 3, pp. 1063–5157, 2007.

[29] B. Q. Minh, F. Pardi, S. Klaere, and A. von Haeseler, "Budgeted phylogenetic diversity on circular split systems," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 6, no. 1, pp. 22–29, 2009.

[30] M. Bordewich and C. Semple, "Nature reserve selection problem: A tight approximation algorithm," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 5, no. 2, pp. 275–280, 2008.

[31] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed.   MIT Press and McGraw-Hill, 2001.

[32] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*.   Belmont, Massachusetts, USA: Athena Scientific, 1997.

[33] L. Underhill, "Optimal and suboptimal reserve selection algorithms," *Biol. Conserv.*, vol. 70, pp. 85–87, 1994.

[34] R. L. Church, D. M. Stoms, and F. W. Davis, "Reserve selection as a maximal covering location problem," *Biol. Conserv.*, vol. 76, pp. 105–112, 1996.

[35] A. S. L. Rodrigues and K. J. Gaston, "Maximising phylogenetic diversity in the selection of networks of conservation areas," *Biol. Conserv.*, vol. 105, no. 1, pp. 103–111, 2002.

[36] R. H. Crozier and R. M. Kusmierski, "Genetic distances and the setting of conservation priorities," in *Conservation Genetics*, V. Loeschcke, J. Tomiuk, and S. K. Jain, Eds. Basel: Birkhauser Verlag, 1994, pp. 227–237.

[37] R. H. Crozier, P.-M. Agapow, and K. Pederson, "Towards complete biodiversity assessment: an evaluation of the subterranean bacterial communities in the oklo region of the sole surviving natural nuclear reactor," *FEMS Microbiol. Ecol.*, vol. 28, pp. 325–334, 1999.

[38] J. B. Kirkpatrick, "An iterative method for establishing priorities for the selection of nature reserves: an example from tasmania," *Biol. Converv.*, vol. 25, pp. 127–134, 1983.

[39] B. Faller, T. Ingram, and C. Semple, "Preserving phylogenetic diversity in ecological networks." 2010, in preparation.