Title
Phylogeography and population genetics of diving beetles in New Guinea

Author
Lam, Athena Wai

Publication Date
2017

Peer reviewed|Thesis/dissertation

Phylogeography and population genetics of diving beetles in New Guinea

By

Athena W Lam

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor George K. Roderick, Chair
Professor Michael Balke
Professor Rosemary G. Gillespie
Professor Rauri C. K. Bowie

Spring 2017

Phylogeography and population genetics of diving beetles in New Guinea

by Athena W Lam

Abstract

Phylogeography and population genetics of diving beetles in New Guinea

Athena W. Lam

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor George K. Roderick, Chair

Predaceous diving beetles (Dytiscidae) are common inhabitants of both lentic and lotic freshwater systems worldwide. They have played a major role in our understanding of the relationship between habitat stability and evolution for dispersal propensity.

Numerous endemic diving beetles can be found in on the island of New Guinea. However, like most micro-fauna on remote islands, their evolutionary history and population ecology is largely unknown. In this series of studies, I use both traditional Sanger sequencing and next-generation sequencing techniques to explore the phylogenetic relationship of beetles in a genus as well as the population genetic patterns within a species.

TABLE OF CONTENT

Acknowledgement

The last 6.5 years involved a lot of hard work, many long nights and weekends, a non-negligible number of failed ventures, and more unexpected twists and turns than I anticipated. But I must say that this journey was thoroughly enjoyable because of the many, many people who came alongside me and supported my scientific endeavors, kept me sane and hopeful, and nourished my intellectual and emotional life.

Many thanks to my mentors at UC Berkeley and The Bavarian State Collection of Zoology in Munich, especially Dr George Roderick and Dr Michael Balke, without whose guidance I would still be spinning in circles trying to figure out how to do good science. I aspire to achieve both scientific greatness and amazing work-life balance as you have.

Many thanks to the past and current members of Evolab at UCB, who were the fabric of my daily life as a graduate student. You are to me much more than my closest colleagues and mentors; I will always see you as family.

Many thanks to my friends and colleagues in Munich, whose friendship has blessed me richly during the most difficult period of this journey, by reminding me that there is more to life than work, by encouraging me to let my hair down and take a swim in the river.

Many thanks to my mom, my very first research sponsors who provided everything a (very) young scientist could dream of in building her early career. I have embarked on, and completed, this fulfilling journey because of your unconditional love, support, and encouragement to pursue my dreams.

Many thanks to Matt Van Dam, my partner in life and all-time favorite scientific collaborator, who never stopped convincing me that I have what it takes to be successful in science and in life. Your excellence is truly inspirational to me. You are a constant reminder that I am not alone, and not an impostor.

I've heard it said that a PhD is a means to an end. I'm not sure what the "end" is, but I certainly know that I will continue to work hard to do good research, spark scientific interest in the students I teach, and inspire appreciation and care for nature in all my audiences.

Thank you all who have been with me on this journey, in all the big and little ways you have contributed.

Introduction

Dispersal is one of the most important life history traits in aquatic insects (Bilton *et al.* 2001), affecting their ability to colonize new habitat, maintain genetic connectivity within a meta-population, and adapt to environmental changes (Bowler & Benton 2005; Hof *et al.* 2012). However, there is an energy tradeoff to the development of dispersal traits, such as the constraints on other reproductive demands (Roff 1986; Denno *et al.* 1989). Dispersal also increase the risk of failure to find the next suitable habitat (Roff 1986; Edwards & Sugg 1993). According to the Habitat Template Concept (HTC), unstable or ephemeral habitats increases an organism's risk of local extinction, therefore, selection favors the evolution of dispersal (Southwood 1977, 1988; Marten *et al.* 2006). Conversely, in an environment that is stable spatially and temporally, selection may favor a less dispersive life history strategy (Roff 1986).

Lotic and lentic habitat differs in their stability and persistence (Ribera & Vogler 2000). Despite some obvious exceptions, running water habitats are in general more stable than small/medium sized standing waterbodies on an evolutionary timescale (Ribera & Vogler 2000). Based on the HTC, while the evolution of dispersal is pertinent for lentic species to avoid local extinction; species are expected to have reduced dispersal ability in stable lotic habitats such as tropical streams.

In **Chapters 1**, I used ddRAD data to explore the population structure of a species of diving beetle (*Exocelina manokorienses*) that is a stream margin (riparian). I found unexpectedly high levels of population subdivision within a 100 km transect. These results are in concordance with predictions of the HTC. My study represents the first genome wide fine-scale population genetic studies on diving beetles and redefines the scale at which the HTC should be considered.

In **Chapter 2**, I explored the population connectivity of different groups of lotic beetles in the *Philaccolilus ameliea* species complex. To my surprise, I revealed complex and idiosyncratic patterns of population structure in different clades. I recover some clusters forming panmictic populations throughout a large geographic range and others lacking genetic exchange between neighboring populations. This suggests that the classical dichromatic correlation between stability and dispersal in freshwater species does not capture the complexity of ecological preferences and their relation to population structuring processes in nature. My results also indicate the formation of cryptic speciation across the complex landscape of New Guinea.

Chapter 3, represents the first phylogenetic reconstruction of *Philaccolilus,* a common and widespread genus of diving beetles in New Guinea. I utilize both species discovery and species validation methods to delimit species of the genus. I also discussed the challenges of delineation of evolutionarily unites in rare and diverse lineages inhabiting remote tropical locations.

References

Bilton DT, Freeland JR, Okamura B (2001) Dispersal in Freshwater Invertebrates. *Annual Review of Ecology and Systematics*, 32, 159–181.

Bowler DE, Benton TG (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological reviews of the Cambridge Philosophical Society*, 80, 205–25.

Denno R, Olmstead K, McCLoud E (1989) Reproductive cost of flight capability: a comparison of life history traits in wing dimorphic planthoppers. *Ecological Entomology*, 14, 31–44.

Edwards JS, Sugg P (1993) Arthropod fallout as a resource in the recolonization of Mount St. Helens. *Ecology*, 74, 954–958.

Hof C, Brändle M, Dehling DM *et al.* (2012) Habitat stability affects dispersal and the ability to track climate change. *Biology Letters*, 8.

Marten A, Brandle M, Brandle R (2006) Habitat type predicts genetic population differentiation in freshwater invertebrates. *Molecular Ecology*, 15, 2643–2651.

Ribera I, Vogler P (2000) Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera. *Biological Journal of the Linnean Society*, 71, 33–52.

Roff D (1986) The evolution of wing dimorphism in insects. *Evolution*, 40, 1009–1020.

Southwood T (1977) Habitat, the templet for ecological strategies? *The Journal of Animal Ecology*, 64, 337–365.

Southwood TRE (1988) Tactics, strategies and templets. *Oikos*, 52, 3–18.

Chapter 1

Phylogeography and population genomics of a widespread Papuan stream margin beetle: testing the Habitat Template Concept across a complex tropical landscape

Abstract
The Habitat Template Concept in freshwater systems contends that lotic species (occurring in habitats along stream courses), are less dispersive than lentic species (occurring in marshlands, puddles, pools, lakes etc.), resulting in populations that are highly structured. Stream courses not only include streaming water, but often tiny stagnant water microhabitats immediately besides the flowing water. Species adapted to such microhabitats are numerous, but remain virtually unstudied. Here, I present data for one of them, the diving beetle *Exocelina manokwariensis* from six localities along a >200 km transect across the Bird's Head Peninsula of New Guinea. I find fine-scale population structure based on mitochondrial CO1 sequences as well as genomic ddRAD sequencing and low levels of connectivity among populations. My results are congruent with most previous phylogenetic and macroecological studies on lotic species and the Habitat Template Concept, and provide the first detailed evaluation of population structure for a lotic species adapted to stagnant water microhabitats besides the flowing water. I argue that fine-scale, taxon focued population genomic studies will help illuminate the effect of a species natural history on their population structure; as well as the geographic extent to which population strcture due to effects of HTC can be observed.

INTRODUCTION

The Habitat Template Concept suggests that the properties and constraints of habitats drive the evolution of ecological traits and evolutionary strategies of its inhabitants (Southwood 1977; Korfiatis & Stamou 1999). In particular, habitat instability is considered one of the most important factors influencing an individual's propensity to disperse (Roff 1994; Bohonak 1999). Species inhabiting ephemeral habitats experience greater risk of local extinction, and therefore, also a greater selective pressure for individuals to disperse (e.g. Roff 1986, 1990, 1994; Denno *et al.* 1996; Travis & Dytham 1999; Grantham 2003; Friedenberg 2003; Marten *et al.* 2006). As a result, populations inhabiting unstable environments are predicted to be less structured as a result of more gene flow than those inhabiting more stable environments (Ribera *et al.* 2001, 2003; Ribera & Vogler 2004).

Freshwater ecosystems are broadly classified as two main habitat types: lotic or running water, such as rivers and streams, and lentic or standing water, such as lakes and ponds. These two habitat types differ in their stability both spatially and temporally, with lotic habitats often considered to be more persistent over geological time (Ribera & Vogler 2000). Although seasonal variation may alter the annual stability of both types of habitats, streams and rivers are typically connected in a drainage network and/or to other water bodies, and thus persist longer in geological time than small to medium sized lakes and ponds that, once dried out, are no longer connected to other aquatic habitats (Hutchinson 1957). Thus, for aquatic organisms, the Habitat Template Concept predicts that an individual's rate of dispersal, and therefore the genetic pattern of its population, is often significantly determined by the flow regime of their habitat; with species in lotic habitats being less dispersive, and thus, exhibiting greater population structure and smaller range size when compared to their lentic counterparts (Ribera 2008). Indeed, several studies show that habitat type is often a good predictor for range size (Ribera & Vogler 2000; Hof *et al.* 2006), dispersal capability (e.g. Ribera & Vogler 2000; Ribera *et al.* 2003; Hof *et al.* 2006; Abellán *et al.* 2009; Damm *et al.* 2010), as well as dispersal-associated traits such as wing size (Arribas *et al.* 2012).

Where molecular genetic data are available greater population structure has generally been associated with lotic species, as found by Marten *et al.* (2006) in a survey of allozyme diversity in 173 species of freshwater invertebrates. Similar results were found by Drotz et al. (2012) for two mayfly species (Ephemeroptera) in Finland based on CO1 data and by Hjalmarsson et al. (2014) for aquatic beetles in Madagascar.

However, despite these results, the link between habitat instability and dispersal is still not fully supported in aquatic systems. Indeed, recent studies suggested that a lentic/lotic dichotomy might be too simplistic and does neither reflect the genuine complexity of aquatic ecosystems nor lineage idiosyncratic dispersal capacity and behaviour. First, within the lentic/lotic classification, microhabitat variation exists and is important. For example, Short & Caterino (2009) examined CO1 data for three lotic species from three different beetle families and found important differences in terms of habitat exploitation and genetic structuring. Indeed, among these three lotic species, only one species (*Eubrianax edwardsii*) occupied the stream bed, while the two other species (*Anacaena signaticollis* and *Stictotarsus striatellus*) respectively occupied detritus at the stream margin or stream pools. Of the three beetles, only *Eubrianax edwardsii* exhibited strong genetic structure. This example highlights the need for more research to understand the importance of the actual species microhabitat along and/or

across the streambed on population structuring. Second, this dichotomic view ignores the importance of an organism's dispersal traits and microhabitat affinity, which are known to play a central role in the degree of gene flow between populations (Alp *et al.* 2012). Most probably, all of these factors interact with one another to form the current population and species distribution patterns (Phillipsen *et al.* 2014).

The rugged and complex New Guinea topography provides the opportunity to study population connectivity of stream organisms in a tropical environment. The island is rich in running water habitats, yet these habitats are separated by mountain ranges and are thus relatively discrete. Despite its high biodiversity and large number of potentially threatened species, the phylogeography and population genetics of the islands' aquatic fauna remain largely understudied (but see McGuigan *et al.* 2000).

Here I study the lotic New Guinea diving beetle species *Exocelina manokwariensis* (Shaverdo *et al.* 2016). These beetles, less than 4mm long, inhabit small patches of stagnant water along low order forest creeks and stream margins. There are about 150 species in New Guinea and most are thought to be narrow endemics (Shaverdo *et al.* 2012). However, a few species have wider ranges, such as *E. manokwariensis*, found across the Bird's Head Peninsula. Using genomic and mitochondrial DNA sequencing I examined this species' population structure with the null hypothesis that inhabiting stable stream margin habitats –although adapted to stagnant water– predicts limited dispersal and strong population structure. six localities along a ~300 km transect were sampled, from altitudes c. 140–1,000 m above sea level as well as from areas with different lithologies and potentially different uplift histories (Table 1) to specifically ask: (1) do we find the expected genetic structuring in these lotic inhabitants, and (2) if so, is the substructuring of the species possibly related to past geological events or position slong the elevational gradient.

MATERIAL AND METHODS

Taxon Sampling and DNA Extraction. We assembled a dataset for 68 specimens from 6 localities (map inlay in Figure 1) (65 had cox1 data, 63 ddRAD data) (Table S1, also see Fig. 1). Genomic DNA was extracted non-destructively from whole beetles using the DNeasy and NucleoSpin 96 Tissue kits (Qiagen, Hilden; Macherey Nagel, Düren, Germany): We first puncture the beetle's metaventrite with a sterile insect pin (size 5), the specimen with exposed tissue is then digested overnight and extracted the next day according to manufacturer's protocol.

Mitochondrial DNA Sequencing. Mitochondrial DNA sequences were obtained for 65 individuals. The 3' end of the mitochondrial cytochrome oxidase subunit 1 (CO1) which is most widely used for molecular taxonomic work in diving beetles (e.g. Monaghan *et al.* 2006) was amplified using the primers Jerry and Pat (Simon *et al.* 1994). PCR reactions were run in a 20 µl total volume containing: 12.5 µL ddH2O, 0.5 µL each of 10 µm primers, 2 µL 10mm DNTPs, 1.25 µL 50mm MgCl2, 5 µL reaction buffer, 0.2 µL Taq polymerase. PCR was performed with an initial denaturation step of 96°C for 3 mins, 35 cycles of 94°C for 30 s, 48°C for 1 min, 72°C for 1 min and a final extension step of 72°C for 10 mins. Sequences were edited using Sequencher 5.0.1 (GeneCodes Corp., Ann Arbor, MI, USA) and aligned in Mesquite 3.04 (Maddison & Maddison 2008) using the ClustalW algorithm (Larkin *et al.* 2007), and were subsequently manually checked for stop codons. Gene alignments were then

concatenated into a combined matrix. These CO1 sequences were deposited in the European Nucleotide Archive (LT615638–LT615714).

ddRAD Sequencing & Output Treatment. A total of 65 individuals were used for ddRAD (double digest restriction site associated DNA sequencing) genotyping. The double digestion was performed with the restriction enzymes SbfI and MseI and we used a modified protocol (see supplementary material S2) from Mastretta-Yanes et al. (2014). In total, we used 32 uniquely barcoded adaptors and 2 Illumina indexes to build the ddRAD library. The library was sequenced once using a single-read protocol on an Illumina 2500 HiSeq (Lausanne Genomics Technology Facility, Lausanne, Switzerland).

To demultiplex, filter and assemble the raw ddRADseq output data, I used pyRAD v.3.0.3 (Eaton 2014). The low quality threshold (Mindepth) was fixed to 6 and the value for Wclust set to 0.88. Additionally, we fixed the maximum sites per read with an error rate > 1% to 4, the minimum number of samples in a final locus (MinCov) to 2 and the maximum proportion of shared polymorphic sites in a locus (MaxSH) set to 3. Other mandatory parameters were kept as defined by default. For the optional parameters, the strictness of filtering (option 21) was fixed at 2, enforcing a strict filter for adaptors, barcodes and cut sites. After having retrieved filtered sequences for each individual, we applied a species-specific cut-off based on coverage to discard individuals with a weak signal. Only one individual was discarded with this additional filter.
Signals of selection were tested for all polymorphic loci using using the Bayesian simulation method of Beaumont & Balding (2004) as implemented in BAYESCAN 2.1 (Foll & Gaggiotti 2008). I used a prior odds value of 10, with 100,000 iterations and a burn-in of 50,000 iterations. We identified loci that were significant outliers at a *q*-value (i.e., false discovery rate) of 0.05.Since most population genetic algorithms assumes neutrality, loci identified as statistical outliers were removed from the dataset before further population analyses.

Population genetics and phylogenetic Inference.
CO1 dataset. To examine the distribution of mtDNA sequence diversity in the six populations, haplotype networks were constructed using the TCS algorithm (Clement *et al.* 2002) implemented in the PopART software (Leigh & Bryant 2015).

I employed molecular dating analyses to infer time of population divergence. I used a relaxed molecular clock in BEAST 2.4.4 (Drummond *et al.* 2012). I used COI codon positions as our character sets and used PartitionFinder version 2.1.1 (Lanfear *et al.* 2012) to find the 'best' partitioning strategy. I used the YULE tree model and pruned each population to the individual with the most complete data. This was done because the YULE model assumes that each terminal is a species/monophyletic coalescing unit. I used the uncorrelated relaxed clock model that has a lognormal distribution of rates, which has been shown to give more accurate estimates than one which assumes an exponential distribution of rates (Baele *et al.* 2013). I set a uniform prior for the mitochondrial clock rate with a mean of 0.034, as these have been reported in dytiscid beetles (Toussaint *et al.* 2014; Tänzler *et al.* 2016). I had four different independent runs of BEAST for $100x10^{6}$ generations. I then examined each run in Tracer and removed the burnin (5%) and combined each run in Logcombiner 1.8.0 (Drummond *et al.* 2012). The maximum credibility tree was generated in TreeAnnotator 1.8.0 (Drummond *et al.* 2012).

ddRAD dataset. I used Maximum-likelihood (ML) based phylogenetic reconstructions for the concatenated dataset (including both informative and uninformative sites). The partitions and corresponding models of substitution were selected using PartitionFinder 2.00 (Lanfear *et al.* 2012) using the 'greedy' algorithm, with character sets grouped by the number of phylogenetically informative sites, under the GTR+G model for RAxML, and the Akaike Information Criterion corrected (AICc) to select among models. The ML analyses were performed using RAxML 8.0.19 (Stamatakis 2014) conducting 20 independent tree searches for the best ML tree. I assessed support for the best ML topology by performing nonparametric bootstrapping using the autoMRE option in RAxML. A calculated bootstrap support value BS ≥ 70 was considered to indicate strong support for a given clade (Hillis & Bull 1993; Erixon *et al.* 2003).

Levels of genetic differentiation between each population pair was estimated by pairwise F_{st} (Weir & Cockerham 1984) using the program GENETIX (Belkhir *et al.* 2004), evaluated using 1,000 permutations. The ddRAD concatenated dataset was used to determine hierarchical levels of genetic structure within and among populations, I conducted a hierarchical analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) using Arlequin 3.5 (Excoffier & Lischer 2010), with significance assessed using 1023 permutations.

In order to identify major genetic clusters in populations of *E. manokwariensis* throughout the Bird's Head Peninsular, I used the Bayesian clustering approach implemented in the program STRUCTURE 2.3.4 (Pritchard *et al.* 2000). I used the SNP dataset including only variable sites. I ran 10 replicates, each using a burn-in length of 100,000 and a run length of 1,000,000 steps, with the admixture and the correlated allele frequencies models without using prior population information (geographic sampling location). I varied the number of clusters (*K)* from 1 to 10.

The broad scale number of clusters was initially determined examining both the posterior probabilities of the data for each K and the ΔK estimator described by Evanno et al. (2005) as calculated in Structure Harvester (Earl & vonHoldt 2012). However, many recent studies suggest that this method tends to underestimate genetic structure in groups with fine-scale genetic structure and in taxa with more complex evolutionary history (e.g. Waples & Gaggiotti 2006; Coulon *et al.* 2008; Levy *et al.* 2012; Lambert *et al.* 2013; Gowen *et al.* 2014). Since I aimed to study fine-scale population structures within a small geographic range, hierarchical approach using the program was subsequently conducted to ran analyses on successively smaller clusters as in Gowen *et al.* (2014). For example, when two distinct population clusters were identified in the full dataset of this study, I ran each of these clusters in its own analysis for $K = 2$ until no geographical clustering was discernable or only one pre-defined population (based on collecting locality) remained. Results from replicates for the inferred K from each run were analyzed in the program CLUMPP (Jakobsson & Rosenberg 2007) to produce averaged matrices of individual and population cluster membership coefficients. Finally, I used the program *distruct* v1.1 (Rosenberg 2004) to produce graphical displays of the resulting barplots.

Principal component analyses (PCA), a multivariate approach, was used as an alternative analytical approach to compare the consistency of the various results obtained from the methods described above. The concatenated ddRAD dataset was used. Specifically, I used GenAlEx 6.1 (Peakall & Smouse 2006) to calculate genetic distances and to convert this into

a covariance matrix with data standardization for the PCA. The first three principal components were plotted in the R package scatterplot3d (Ligges & Martin 2003).

Structuring can result from limited gene flow amoung populations within an area. This phenomenon is known as isolation-by-distance (IBD) and is a simple consequence of limited dispersal across space. IBD was evaluated by testing for a statistically significant association between genetic differentiation as measured by $F_{st}/(1 - F_{st})$ (Rousset 1997) and log-transformed geographic distance among each of the populations. The significance of the correlation was evaluated with a Mantel test (Mantel 1967) using 10,000 permutations as implemented in GenAlEx 6.1.

RESULTS

Mitochondrial and genomic data.

The trimmed CO1 alignment comprised of 729 base pairs for 65 individuals. For the ddRAD SEQ, the Hiseq sequencing run produced 258 million reads. After pyRAD filtering, de novo assembly and implementation of the minimum coverage cut off, I recovered 3,196 loci encompassing 2,689 SNPs for 63 individuals. BAYESCAN determined that no loci displayed signal of selection at $q = 0.05$ (Fig S3), therefore all SNPs were retained for population genetic analyses.

Population and phylogenetic inference.

CO1 dataset. The TCS network shows that the CO1 haplotypes are broadly divided in two groups: (1) a "Western cluster", consisting of populations from BH041 and 044; and (2) an "Eastern cluster" consisting of populations form BH028, 033, 034, and 039. Individuals from BH033 and 034 all share a single haplotype. Though interconnected beyond the 95% cut off, the Western and Eastern clades remain distinct (Fig.1D). Our dated phylogeny shows that the split at the root node between the outgroup (*E. anggiensis*) and *E. manokwariensis* occurred at 1.73 Ma (95% HPD:1.15–2.35 Ma). The split at the node at the base of *E. manokwariensis* that separate the Eastern and Western clusters occurred at 0.77 Ma (95% HPD:0.44–1.13 Ma). The split between the BH033,034 clade and BH028/BH039 clade occurred at 0.39 Ma (95% HPD:0.15–0.59 Ma). The remainder of the terminal splits occurred between 0.06–0.18 Ma (Fig 2), suggesting relatively recent divergence of *E. manokwariensis* populations and subsequent isolation.

ddRAD dataset. The RAxML analyses resulted in a topology with a highly supported backbone (Fig. 1C). The phylogeny is again split into the "Eastern clade" and the "Western clade" as described above, both clades being monophyletic. In the "Eastern clade" BH028 is sister to the 2 subclades - BH039 and BH033, BH034. In the "Western clade" BH044 and BH041 are sister to each other. Each of these subclade remains monophyletic.

The AMOVA revealed moderate amount of variation among populations in the Bird's Head Peninsular (54.07%, $P = 0.00$) (Table 2). These patterns were congruent with those from pairwise F_{st} estimates (Table 3), with moderate to great values (Hartl & Clark 1997) found between all populations except BH033 and BH034, which are located extremely close to each other geographically. The F_{st} values among populations clustered within the Eastern and Western clades, respectively, are generally large ($F_{st} = 0.18–0.26$). Values within each clade

are generally moderate (F_{st} = 0.05–0.15) with the exception of BH033 and BH034 with negligible differentiation (F_{st} = 0.02).

The PCA analysis showed that the first three components explained a majority of the total variance (62.22% cumulative). Overall the PCA analysis recovered similar major patterns as the other analyses: 2 separate clusters can be observed representing the Eastern and Western clades described above. Each population, except BH033 and BH034, remain somewhat distinct (Fig. 3).

In the initial Structure run of K = 1–10 using the full dataset, the ΔK test (Evanno *et al.* 2005) depicted K = 2 as the best model for the data (Fig. 1B (i); S4 for ΔK table and graph). As with other analyses, the two major and most distinct clades are: The Eastern Clade including BH028, BH033, BH034, BH039 and the Western Clade including BH041 & BH044. However, subsequent hierarchical analyses revealed strong structure within these clades that was highly justified based on biology and geography. Indeed, Structure run on successively smaller clusters revealed five distinct geographic clusters that were largely uniform in their population assignment: The Western clade was split into two distinct population with no significant admixture, i.e. (1) BH041 and (2) BH044 (Fig 1B (ii)). In the Western clade, (3) BH028 first split off from the rest of the group and seems to be a distinct population from the rest of the clade (Fig. 1B (iii)); (4) BH039 also forms a different population from BH033 & BH034, (Fig. 1B (iv)) but Structure results suggest a moderate level of admixture between the two groups; finally, (5) BH033 and BH034 form a single population (Fig. 1B (v)).

The Mantel test showed no significant correlation between genetic distance and geographic distance (R^2 = 0.161, P = 0.087) (Fig 4).

DISCUSSION

Here, I examined the genetic structure of a comparably widespread lotic beetle species across the Bird's Head Peninsula of New Guinea. The microhabitat of this species are small patches of completely stagnant water (edge of backflows, water filled holes in bedrock, tiny waterholes on gravelly stream edge) along the course of forest creeks, often immediately besides the flowing water. Since these habitats will be flooded with every rainfall, the beetles might avoid drift by hiding in the interstitial where they will usually dig in when disturbed by a collector. Congruent with most previous studies testing the Habitat Template Concept in freshwater systems, I found pronounced fine scale genetic structure with limited gene flow among populations. In addition to mtDNA CO1 sequence data, I utilized next-generation sequencing techniques to obtain ample markers throughout the genome. RADseq data gives our results added robustness and allowed us to more accurately quantify the level of genetic structure. In fact, in doing so I can reveal a much clearer isolation of populations and very pronounced geographic structure than by relying merely on CO1 data.

While *Exocelina manokwariensis* as defined morphologically (Shaverdo *et al.* 2016) has a comparably wide range when compared to other *Exocelina* species, our analyses reveal pronounced genetic structure. Excluding the populations BH033 and BH034 (which are less than 1 km from each other on the same mountain slope), average pairwise distance between differentiated populations is only 45 km. This suggests a low level of dispersal, in agreement with the Habitat Template Concept and previous studies that found correlations between

habitat type and dispersal in aquatic insects. Since patterns of population subdivisions are found in both mitochondrial and genomic markers, there is no apparent sex-biased dispersal in this species, as one would expect as the sexes are not sexually dimorphic, i.e. similar body size and both having flight wings.

Short and Caterino (2009) however found that levels of genetic structure can vary widely even for species of lotic inhabitants found in overlapping areas within Southern California. In the three species studied, strong genetic structure was only observed in a species adapted to the rapid flowing water, while populations of the stream margin and streampool inhabitants had less or lack genetic structure. The authors attribute these differences to the beetles' different microhabitat preferences and specificities, as well as other natural history traits. In this respect, our results were contradictory, as the results of Short & Caterino (2009) would less population structure in species inhabiting lotic habitats without water current. In another comparative population genetic study, Hjalmarsson et al. (2014) also included a genus of tropical diving beetles, *Madaglymbus*, that are at least partially specialized to stream margins. Similarly, they found a significant level of population structure, based on CO1, for nine species of diving beetles in this genus. However, an increase in dispersal was not observed by the stream margin specialist in our study. These results all suggest that even microhabitat mights not always be a predictor of a species' population genetic pattern, which is thus highly lineage indiosyncratic.

The high level of genetic distance without signal of IBD found in our study suggests that gene flow is low and random genetic drift is the main mechanism driving the divergence of populations in *E. manokwariensis* (Hutchinson 1957; Phillipsen *et al.* 2014). As a result, each population is evolving independently making them more likely to become microendemic.

The geomorphology of the Bird's Head Peninsula that I recognize today, only began to take shape in the past few million years (~< 3 Ma) (e.g. Haq & Al-Qahtani 2005; Snedden & Liu 2010). Our tentatively dated phylogeny suggests that *E. manokwariensis* is a recent species and its presence on different geological elements might be the result of relatively recent dispersal across an existing complex landscape. However, there is limited current gene flow. Such a pattern would be expected to occur prior to the formation of microendemic species, of which the New Guinea fauna is extraordinarily rich. Factors that prevent species from continued range expansion would therefore be intrinsic and the observed pattern not directly linked to the geological evolution of the area. I suggest that our experiment should be extended to study such biotic factors, as well as investigate different regions of Bird's Head that have evolved in more pronounced geographic isolation, testing for possible interactions of geology and biotic evolution.

References

Abellán P, Millán A, Ribera I (2009) Parallel habitat-driven differences in the phylogeographical structure of two independent lineages of Mediterranean saline water beetles. *Molecular ecology*, **18**, 3885–902.

Alp M, Keller I, Westram AM, Robinson CT (2012) How river structure and biological traits influence gene flow: A population genetic study of two stream invertebrates with differing dispersal abilities. *Freshwater Biology*, **57**, 969–981.

Arribas P, Velasco J, Abellán P *et al*. (2012) Dispersal ability rather than ecological tolerance drives differences in range size between lentic and lotic water beetles (Coleoptera: Hydrophilidae). *Journal of Biogeography*, **39**, 984–994.

Bailly V, Pubellier M, Ringenbach J-C, de Sigoyer J, Sapin F (2009) Deformation zone "jumps" in a young convergent setting; the Lengguru fold-and-thrust belt, New Guinea Island. *Lithos*, **113**, 306–317.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, Population genetics software for Windows TM.

Bohonak A (1999) Dispersal, gene flow, and population structure. *Quarterly review of biology*, 21–45.

Clement M, Snell Q, Walker P, Posada D, Crandall K (2002) TCS: Estimating gene genealogies. *ipdps*.

Coulon A, Fitzpatrick JW, Bowman R *et al*. (2008) Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (Aphelocoma coerulescens). *Molecular Ecology*, **17**, 1685–1701.

Damm S, Dijkstra K-DB, Hadrys H (2010) Red drifters and dark residents: The phylogeny and ecology of a Plio-Pleistocene dragonfly radiation reflects Africa's changing environment (Odonata, Libellulidae, Trithemis). *Molecular Phylogenetics and Evolution*, **54**, 870–882.

Denno RF, Roderick GK, Peterson MA *et al*. (1996) Habitat persistence underlies intraspecific variation in the dispersal strategies of planthoppers. *Ecological Monographs*, **66**, 389–408.

Drotz MK, Savolainen E, Saura A, Ståhls G (2012) The genetic population structure of lotic and lentic mayflies of the Baetis vernus group (Ephemeroptera: Baetidae). *The Canadian Entomologist*, **144**, 679–690.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, **29**, 1969–73.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic biology*, **52**, 665–673.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplo- types: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.

François C, de Sigoyer J, Pubellier M *et al*. (2016) Short-lived subduction and exhumation in Western Papua (Wandamen peninsula): Co-existence of HP and HT metamorphic rocks in a young geodynamic setting. *Lithos*, **266**, 44–63.
Friedenberg NA (2003) Experimental evolution of dispersal in spatiotemporally variable microcosms. *Ecology Letters*, **6**, 953–959.
Gowen FC, Maley JM, Cicero C *et al*. (2014) Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC evolutionary biology*, **14**, 135.
Grantham BA (2003) Dispersal potential of marine invertebrates in diverse habitats. *Ecological Applications*, **13**, 108.
Haq BU, Al-Qahtani AM (2005) Phanerozoic cycle of sea -level change on the Arabian Platform. *GeoArabia*, **10**, 127–160.
Harrison SSC (2000) The importance of aquatic margins to invertebrates in English chalk streams. *Archiv fur Hydrobiologie*, **149**, 213–240.
Hartl DL, Clark AG (1997) *Principles of population genetics*. Sinauer Associates.
Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**, 182–192.
Hjalmarsson A, Bergsten J, Monaghan M (2014) Dispersal is linked to habitat use in 59 species of water beetles (Coleoptera: Adephaga) on Madagascar. *Ecography*, **38.7**, 732–739.
Hof C, Brändle M, Brandl R (2006) Lentic odonates have larger and more northern ranges than lotic species. *Journal of Biogeography*, **33.1**, 63–70.
Hutchinson GE (1957) Treatise on Limnology. 3V. V1-Geography Physics and Chemistry. V2-Introduction to Lake Biology and Limnoplankton. V3-Limnological Botany. *John Wiley & Sons*.
Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
Korfiatis KJ, Stamou GP (1999) Habitat Templets and the Changing Worldview of Ecology. *Biology and Philosophy*, **14**, 375–393.
Lambert SM, Geneva AJ, Luke Mahler D, Glor RE (2013) Using genomic data to revisit an early example of reproductive character displacement in Haitian Anolis lizards. *Molecular Ecology*, **22**, 3981–3995.
Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, **29**, 1695–701.
Larkin MA, Blackshields G, Brown NP *et al*. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
Leigh JW, Bryant D (2015) Popart: Full-Feature Software for Haplotype Network Construction (S Nakagawa, Ed,). *Methods in Ecology and Evolution*, **6**, 1110–1116.
Levy E, Kennington WJ, Tomkins JL, LeBas NR (2012) Phylogeography and Population Genetic Structure of the Ornate Dragon Lizard, Ctenophorus ornatus. *PLoS ONE*, **7**, 853–858.
Ligges U, Martin M (2003) Scatterplot3d – an R package for Visualizing Multivariate Data. *Cran*, **8**, 1–36.
Mantel N (1967) Cancer Research. *Nature*, **214**, 637–637.
Marten A, Brandle M, Brandle R (2006) Habitat type predicts genetic population differentiation in freshwater invertebrates. *Molecular Ecology*, **15**, 2643–2651.
Mastretta-Yanes A, Arrigo N, Alvarez N *et al*. (2014) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.
McGuigan K, Zhu D, Allen GR, Moritz C (2000) Phylogenetic relationships and historical biogeography of melanotaeniid fishes in Australia and New Guinea. *Marine and Freshwater Research*, **51**, 535–541.

Monaghan MT, Balke M, Pons J, Vogler AP (2006) Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 887–893.
Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
Phillipsen IC, Kirk EH, Bogan MT *et al*. (2014) Dispersal ability and habitat requirements determine landscape-level genetic patterns in desert aquatic insects. *Molecular Ecology*, **24**, 54–69.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Ribera I (2008) Habitat constraints and the generation of diversity in freshwater macroinvertebrates. In: *Aquatic insects: challenges to populations* (eds Lancaster J, Briers RA), pp. 289–311. CABI, Wallingford.
Ribera I, Dolédec S, Downie I, Foster G (2001) Effect of land disturbance and stress on species traits of ground beetle assemblages. *Ecology*, **82**, 1112–1129.
Ribera I, Foster G, Vogler A (2003) Does habitat use explain large scale species richness patterns of aquatic beetles in Europe? *Ecography*, **26**, 145–152.
Ribera I, Vogler P (2000) Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera. *Biological Journal of the Linnean Society*, **71**, 33–52.
Ribera I, Vogler A (2004) Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, **13**, 179–193.
Roff D (1986) The evolution of wing dimorphism in insects. *Evolution*, **40**, 1009–1020.
Roff DA (1990) The evolution of flightlessness in insects. *Ecological Monographs*, **60**, 389–421.
Roff D (1994) Habitat persistence and the evolution of wing dimorphism in insects. *American Naturalist*, **144**, 772–798.
Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–28.
Shaverdo H, Panjaitan R, Balke M (2016) A new, widely distributed species of the *Exocelina ekari*-group from West Papua (Coleoptera, Dytiscidae, Copelatinae). *ZooKeys*, **2016**, 69–85.
Shaverdo H V., Surbakti S, Hendrich L, Balke M (2012) Introduction of the *Exocelina ekari*-group with descriptions of 22 new species from New Guinea (Coleoptera, Dytiscidae, Copelatinae). *ZooKeys*, **250**, 1–76.
Short AEZ, Caterino MS (2009) On the validity of habitat as a predictor of genetic structure in aquatic systems: A comparative study using California water beetles. *Molecular Ecology*, **18**, 403–414.
Simon C, Frati F, Beckenbach A *et al*. (1994) Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. *Annals of the Entomological Society of America*, **87**, 651–701.
Snedden JW, Liu C (2010) A Compilation of Phanerozoic Sea-Level Change, Coastal Onlaps and Recommended Sequence Designations. *AAPG Search and Discovery article 40594*.
Southwood T (1977) Habitat, the templet for ecological strategies? *The Journal of Animal Ecology*, **64**, 337–365.
Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
Stanley EH, Fisher SG, Grimm NB (1997) Ecosystem Expansion and Contraction in Streams. *BioScience*, **47**, 427–435.
Tänzler R, Van Dam MH, Toussaint EFA *et al*. (2016) Macroevolution of hyperdiverse flightless beetles reflects the complex geological history of the Sunda Arc. *Scientific Reports*, **6**, 18793.
Toussaint EFA, Hall R, Monaghan MT *et al*. (2014) The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature communications*, **5**, 4001.
Travis JMJ, Dytham C (1999) Habitat persistence, habitat availability and the evolution of dispersal. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 723–728.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.
Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.

Data Accessibility
DNA sequences: European Nucleotide Archive (LT615638–LT615714).

Sample	Latitude	Longitude	Geological Formation	Geological age and description	Estimated time of uplift
BH028	-0.88385	132.73706	Kemum Formation	Paleozoic metasedimentary rocks overlain by Miocene (16-15 Ma) limestone and Oligocene sandstone	Uplift sometime between ~15 Ma and 0 Ma
BH033	-0.78320	133.07214	Quaternary Alluvium	Quaternary alluvium deposited on top of the Triassic Netoni granitoids	Uplift sometime between 3 Ma and 0 Ma
BH034	-0.77452	133.06993	Quaternary Alluvium	Quaternary alluvium deposited on top of the Triassic Netoni granitoids	Uplift sometime between 3 Ma and 0 Ma
BH039	-0.90758	133.92147	Maruni Limestone /Quaternary Alluvium	Boundary between the Lower-Middle Miocene Maruni Limestone and Quaternary alluvium	Uplift sometime between ~15 Ma and 0 Ma
BH041	-0.76297	131.61770	Limestone within the Sorong Fault Zone	Miocene limestones within the Sorong Fault Zone	Uplift sometime between ~23 Ma and 0 Ma
BH044	-0.69750	132.07225	Tamrau Formation	Middle Jurassic to Cretaceous metasedimentary rocks unconformably overlain by Pliocene melange and Quaternary alluvial deposits	Uplift sometime between ~66 Ma and 0 Ma

Table 1: Geological information and likely uplift history that corresponds with each of the sampling sites according to existing geological maps (Pieters et a., 1989; Amri et al., 1990; Robinson et al., 1990) as well as inferences drawn from recently collected, but unpublished field data

Source of variation	df	Sum of squares	Variance components	Percentage of variation	Fixation index
Among populations	5	856.448	15.34367 Va	54.0703	FST=0.54070, P=0.000000
Within population	62	656.572	13.03359 Vb	45.9297	
Total		1513.02	28.37726		

Table 2: Analyses of molecular variance (AMOVA) of the six populations in the Bird's head peninsular using ddRAD data

Location	BH028	BH033	BH034	BH039	BH041	BH044
BH028	-	**0.15***	**0.14***	**0.13***	**0.18***	**0.15***
BH033		-	**0.02**	**0.05***	**0.17***	**0.26***
BH034			-	**0.05***	**0.15***	**0.20***
BH039				-	**0.15***	**0.18***
BH041					-	**0.13***

Table 3: Pairwise Fst estimates based on ddRAD data. Values with * were significant (P<0.05). Pairwise values for among clade comparisons are shaded

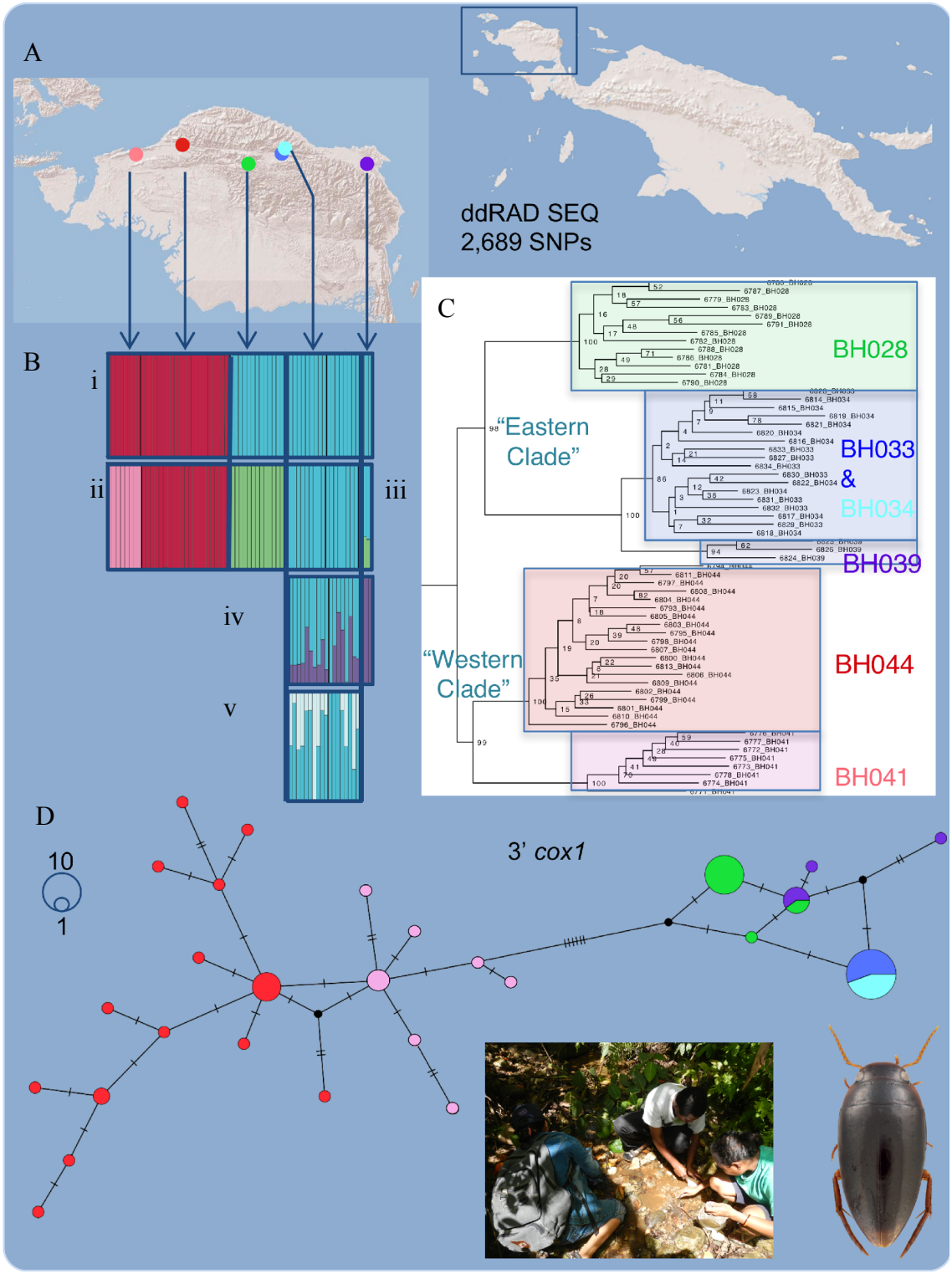


Fig 1: Population genetics of *Exocelina manokwariensis*. **A** Bird's Head peninsular with collecting localities of the six populations studied, featured in map of New Guinea in upper right (locality colors correspond to colors in Figs. C and D). **B** Bayesian clustering analyses of ddRAD SNPs data in STRUCTURE. Barplots from STRUCTURE runs for successively smaller genetic clusters (*K=1-10* for the first run, *K* = 2 for each successive run). (Bi) The first run split the individuals broadly into the Eastern Clade and Western Clade. (Bii) The Western Clade split into two distinct groups corresponding to the two collecting localities,

BH041, BH044. (Biii) Population BH028 is split from the rest of the Eastern Clade. (Biv) There are some evidence for differential assignment between BH039 and BH033 + BH034. (Bv) Individuals form BH033 and BH034 have mixed assignments and showed no geographic structure. **C**Maximum Likelihood (RAxML) tree based on ddRAD dataset with bootstrap support values. Numbers and colors correspond to collecting locations. **D** TCS network based on CO1 sequences. Colors correspond to locality colors on map. Underneath, habitus of beetle (Foto: Harald Schillhammer, Vienna).

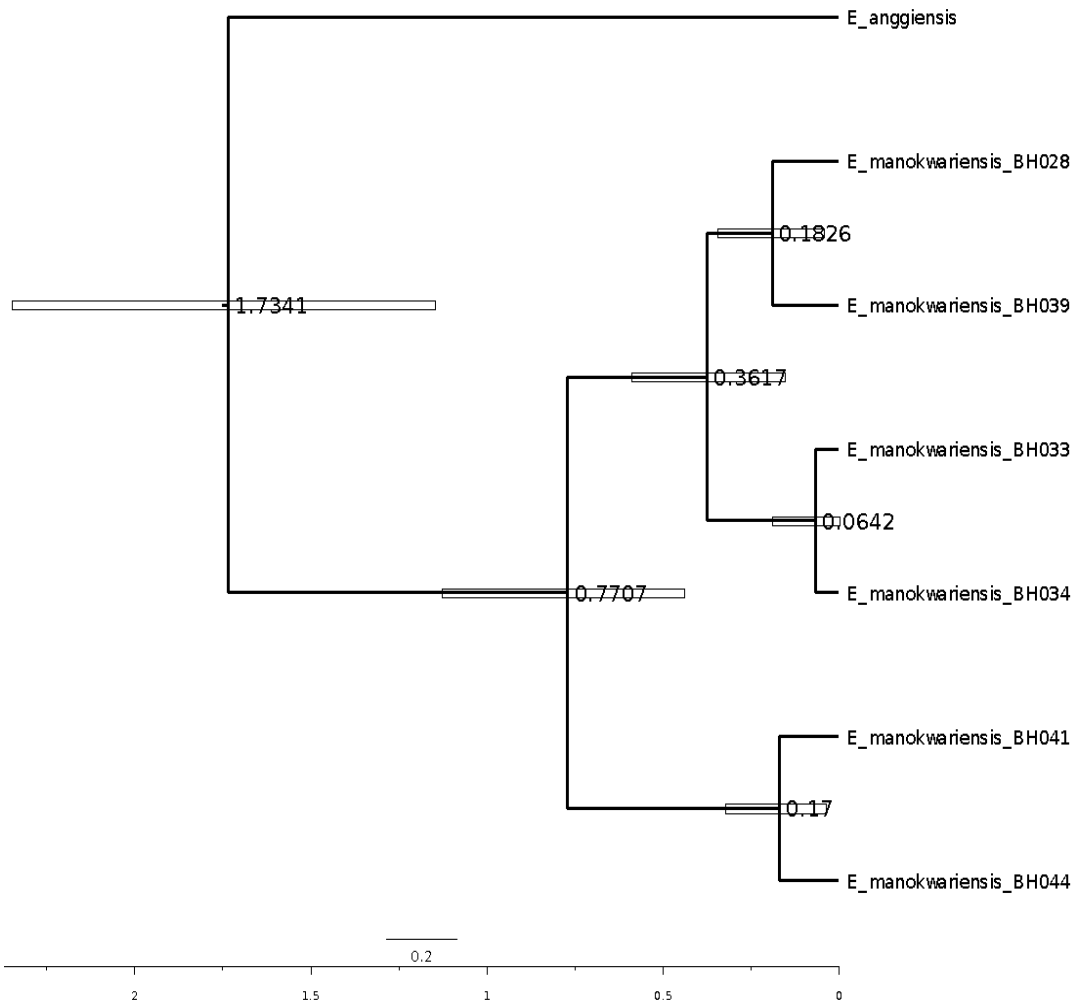


Fig 2: Dated phylogeny of *Exocelina manokwariensis* based on CO1 sequences

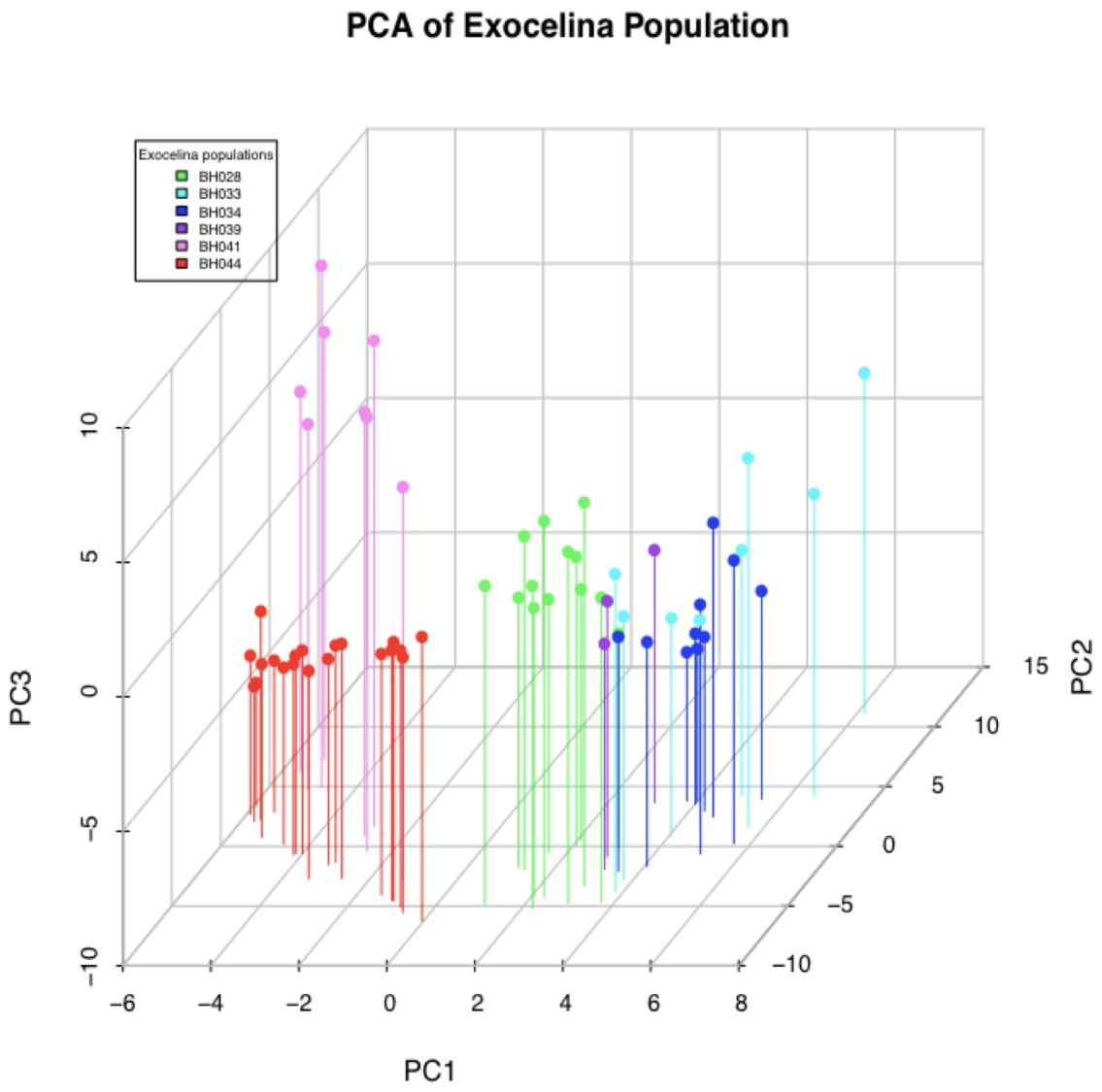


Fig 3: Three-dimensional plot of a Principal Coordinates Analysis based on individual ddRAD genotypes. Individuals are color-coded according to collection locality.

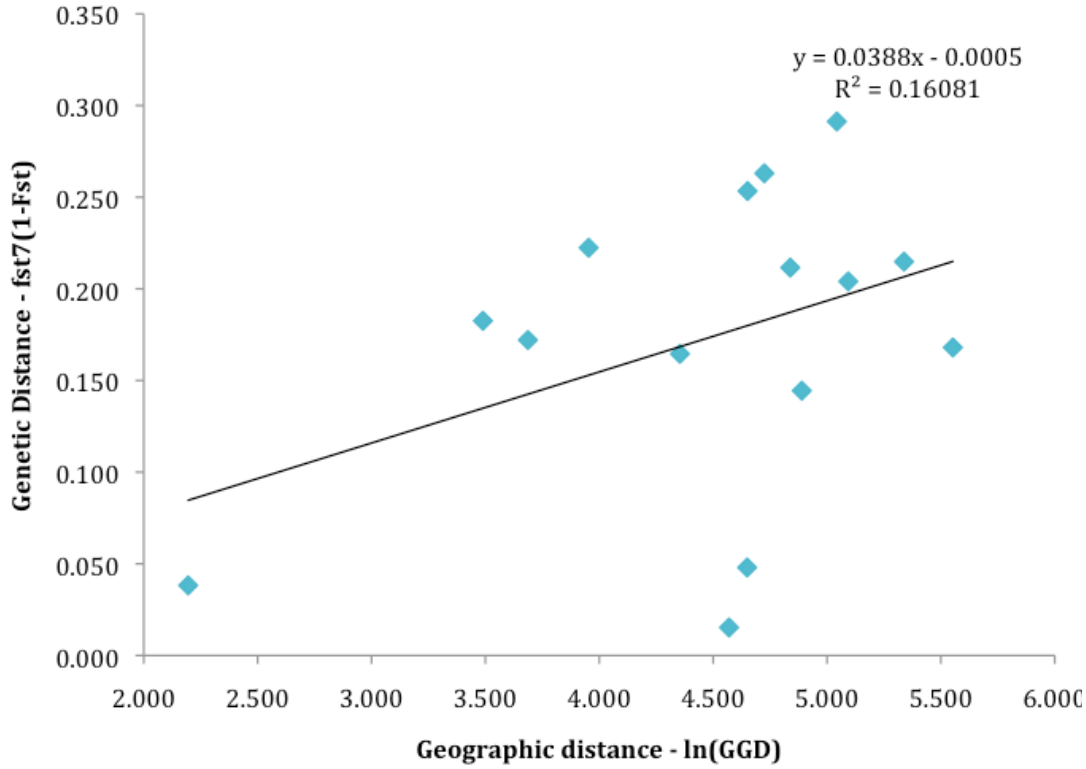


Fig 4: Relationships between genetic distances and geographic distances between pairs of populations. Genetic distances were based on the ddRAD sequences calculated on linearized Fst. Geographic distances are log-transformed Euclidean distances in km.

SUPPLEMENTARY MATERIAL

S1 *Exocelina manokwariensis* specimens used in this study

ddRAD	Voucher ID	Location ID	Country	Province	Locality	Elevation
x	MB 6779	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6780	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6781	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6782	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6783	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6784	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6785	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6786	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6787	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6788	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6789	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6790	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
x	MB 6791	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
	MB 6792	BH028	Indonesia	Papua Barat	Fumato to Kebar, forest stream	674m
	MB 6496	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
	MB 6497	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6771	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6772	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6773	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6774	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6775	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6776	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6777	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
x	MB 6778	BH041	Indonesia	Papua Barat	Sorong-Sausapor	300m
	MB 6498	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
	MB 6499	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6793	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6794	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6795	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6796	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6797	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6798	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6799	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6800	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6801	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6802	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6803	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6804	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6805	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6806	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6807	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6808	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6809	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6810	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6811	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6812	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6813	BH044	Indonesia	Papua Barat	Sausapor-Fef	157m
x	MB 6814	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6815	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6816	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6817	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6818	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6819	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6820	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6821	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6822	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6823	BH034	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream, puddles	1050m
x	MB 6824	BH039	Indonesia	Papua Barat	Manokwari, Maripi, creek white pebbles	135m

x	MB 6825	BH039	Indonesia	Papua Barat	Manokwari, Maripi, creek white pebbles	135m
x	MB 6826	BH039	Indonesia	Papua Barat	Manokwari, Maripi, creek white pebbles	135m
x	MB 6827	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m
x	MB 6828	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m
x	MB 6829	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m
x	MB 6830	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m
x	MB 6831	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m
x	MB 6832	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m
x	MB 6833	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m
x	MB 6834	BH033	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream	750m

S2: Double digest Restriction site associated DNA (RAD) sequencing protocol

This protocol is a modified version of A. Mastretta-Yanes (2014 *Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference)*
Summary of main modifications:

- Q5 high fidelity polymerase used instead of phusion Taq
- Size selection using blue Pippin instead of agarose gel extraction
- Purification and equi-molar pool of PCR products.

1. **Double restriction digest**

a. Prepare master mix I, mix and centrifuge.

MASTER MIX I: DIGESTION	1x (μL)
CutSmart buffer 10x	0.90
H2O	1.90
MseI (10,000 U/ml)	0.10
SbfI (HF) (20,000 U/ml)	0.10

b. Place 6 μL of sample DNA in each well of a plate.
c. Add 3 μL of the master mix I to each well.
d. Cover and seal the plate, centrifuge and incubate at 37°C for 3 hours* on a thermal cycler with a heated lid. Heat kill the enzyme with 20 mins at 65ºC. Keep at 4ºC afterwards.

2. Adaptor Ligation

a. Thaw P1 and P2 adaptors. These adaptors should already be annealed (see Mastretta-Yanes et al. 2014 for adaptor annealing).
b. Prepare master mix II, mix and centrifuge.

MASTER MIX II: LIGATION	1x (μL)
CutSmart buffer 10x	0.26
100 mM ATP	0.12
H2O	0.06
P2 (MseI) adapter 10 uM	1.00

T4 DNA Ligase (400,000 U/ml)	0.17

c. Add 1.6 μL of the master mix II to each well of the restriction digested DNA.
d. Add 1 μL of the P1 (SbfI) adaptor to each well (a unique barcoded adaptor for each DNA sample)
e. The total reaction volume should now be 11.6 μL. Cover and seal the plate, vortex softly, centrifuge and incubate at 16° C for 3 hours on a thermocycler.
f. Dilute the Restriction-Ligation reaction with 40 μL of Tris 10 mM.

3. Purification

a. *AMPure* purification with ratio of 0.8x. Resuspend in 40 μL of Tris 10 mM

4. PCR Amplification

In order to minimize amplification bias between samples, two individual PCRs were performed.
a. Prepare master mix III, vortex and centrifuge.
In order to maintain color balance for each base of the index read being sequenced, Illumina recommends to use the following combinations; if only 2 index primers will be used use the ILLPCR2_ind06 and ILLPCR2_ind12, if three primers, use 4, 6, 12. If six primers use 2,4,5,6,7,12.

MASTER MIX III: PCR	1x (μL)
H2O	2.15
Q5 Buffer	2
dNTP (25mM)	0.08
PCR Primer Mix	0.67
GC enhancer	2
Q5 Taq	0.1

b. Add 7 μL of the combined master mix III to each well of a plate.
c. Add 3 μL of the diluted ligation product from step II or of the purification product if step III was done.
d. Thermal cycler profile for this PCR: 98° C for 30s; 20 cycles of: 98° C for 20s, 60° C for 30s, 72° C for 40s; final extension at 72° C for 10 min.
e. Prepare master mix IV

MASTER MIX IV: PCR final cycle	1x (μL)
H2O	0.05
Q5 Buffer	0.2

PCR primer mix	0.67
dNTP (25 mM)	0.08

f. Add 1 μL to each PCR product, run thermocycler profile as follows: 98° C for 3 min, 60° C for 2 min, 72° C for 12 min.

5. Pool replicates and verify amplification success

a. Pool the two replicates together (final volume of 22 ul/well)

b. For each pool verify amplification success through electrophoresis

6. AMPure purification and pool samples per plate

a. Purify each well of the plates using *AMPure* ratio 1x. Re-elute in 40 ul EB.

b. Accordingly, to the electrophoresis profile pool samples together:

I. If there are no obvious disparity in terms of amplification between the samples of a same plate, pool all wells together

II. If obvious disparity

i. Divide samples in classes according to their amplification success

ii. Select 5 samples per class, quantify using Qbit and calculate mean concentrations

iii. Pool according the concentration differences among classes.

7. Quantify individual index pools, standardize concentrations and pool indexes

a. Quantify concentrations of each pooled index using *Qbit*

b. Bring pools to equal molarity

c. Pool half (in order to keep a backup) of each different indexes pools together

d. Re-concentrate with *SPEEDVAX*

8. Size selection

a. Analyze library profile with a Fragment Analyzer in order to determine the optimal size to select

b. Size select using *Blue Pippin 3%*

c. Verify size selection success with FA

9. Sequence with Illumina Hiseq single end (100bp)

Table 1. Oligo sequences for PCR primers used and MseI adaptors (5'-3')xxxxxxxxxxxxxxx

ID	Sequence	Index sequence **
ILLPCR2_ind 06	C*A*A GCA GAA GAC GGC ATA CGA GAT **CAC TGT** GTG ACT GGA GTT CAG ACG TGT GC	ACAGTG
ILLPCR2_ind 12	C*A*A GCA GAA GAC GGC ATA CGA GAT **GTA GCC** GTG ACT GGA GTT CAG ACG TGT GC	GGCTAC
ILLPCR1	A*A*T GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T	
P2.1 MseI	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T	

P2.2 MseI	/5Phos/TAA GAT CGG AAG AGC GAG AAC AA	

Table 2. Specific P1 oligo sequences for SbfI adapters (5'-3').

1	ACGG	P1_01.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTACGGTGCA	P1_01.2	CCGTAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
2	TGCT	P1_02.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTTGCTTGCA	P1_02.2	AGCAAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
3	CATA	P1_03.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTCATATGCA	P1_03.2	TATGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
4	CGAG	P1_04.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTCGAGTGCA	P1_04.2	CTCGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
5	GCTT	P1_05.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTGCTTTGCA	P1_05.2	AAGCAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
6	ATCA	P1_06.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTATCATGCA	P1_06.2	TGATAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
7	GACG	P1_07.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTGACGTGCA	P1_07.2	CGTCAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
8	CTGT	P1_08.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTCTGTTGCA	P1_08.2	ACAGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
9	TCAA	P1_09.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTTCAATGCA	P1_09.2	TTGAAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
10	AGTCA	P1_10.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTAGTCATGCA	P1_10.2	TGACTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT
11	TCACG	P1_11.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTTCACGTGCA	P1_11.2	CGTGAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT
12	CTGCA	P1_12.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTCTGCATGCA	P1_12.2	TGCAGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT
13	CATCG	P1_13.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTCATCGTGCA	P1_13.2	CGATGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT
14	ATCGA	P1_14.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTATCGATGCA	P1_14.2	TCGATAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT
15	TCGAA	P1_15.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTTCGAATGCA	P1_15.2	TTCGAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT
16	ACCTG	P1_16.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTACCTGTGCA	P1_16.2	CAGGTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT
17	CTCAG	P1_17.1	ACACTCTTTCCCTACACGACGCT CTTCCGATCTCTCAGTGCA	P1_17.2	CTGAGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGT GT

18	CGCTA	P1_18.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTATGCA	P1_18.2	TAGCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
19	CCTGA	P1_19.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTGATGCA	P1_19.2	TCAGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
20	CGACT	P1_20.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACTTGCA	P1_20.2	AGTCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
21	ACGCT	P1_21.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGCTTGCA	P1_21.2	AGCGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
22	GCCAT	P1_22.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCATTGCA	P1_22.2	ATGGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
23	CACGT	P1_23.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACGTTGCA	P1_23.2	ACGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
24	GTTCCA	P1_24.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCCATGCA	P1_24.2	TGGAACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
25	TGTGCA	P1_25.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTGCATGCA	P1_25.2	TGCACAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
26	TTGACA	P1_26.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGACATGCA	P1_26.2	TGTCAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
27	AGCTGA	P1_27.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCTGATGCA	P1_27.2	TCAGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
28	TGGCAA	P1_28.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCAATGCA	P1_28.2	TTGCCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
29	CTATCG	P1_29.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTATCGTGCA	P1_29.2	CGATAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
30	GCTGAA	P1_30.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTGAATGCA	P1_30.2	TTCAGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
31	TTCCGA	P1_31.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTCCGATGCA	P1_31.2	TCGGAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
32	GACTCT	P1_32.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACTCTTGCA	P1_32.2	AGAGTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
33	ATGGCG	P1_33.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGGCGTGCA	P1_33.2	CGCCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
34	TCATGG	P1_34.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCATGGTGCA	P1_34.2	CCATGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

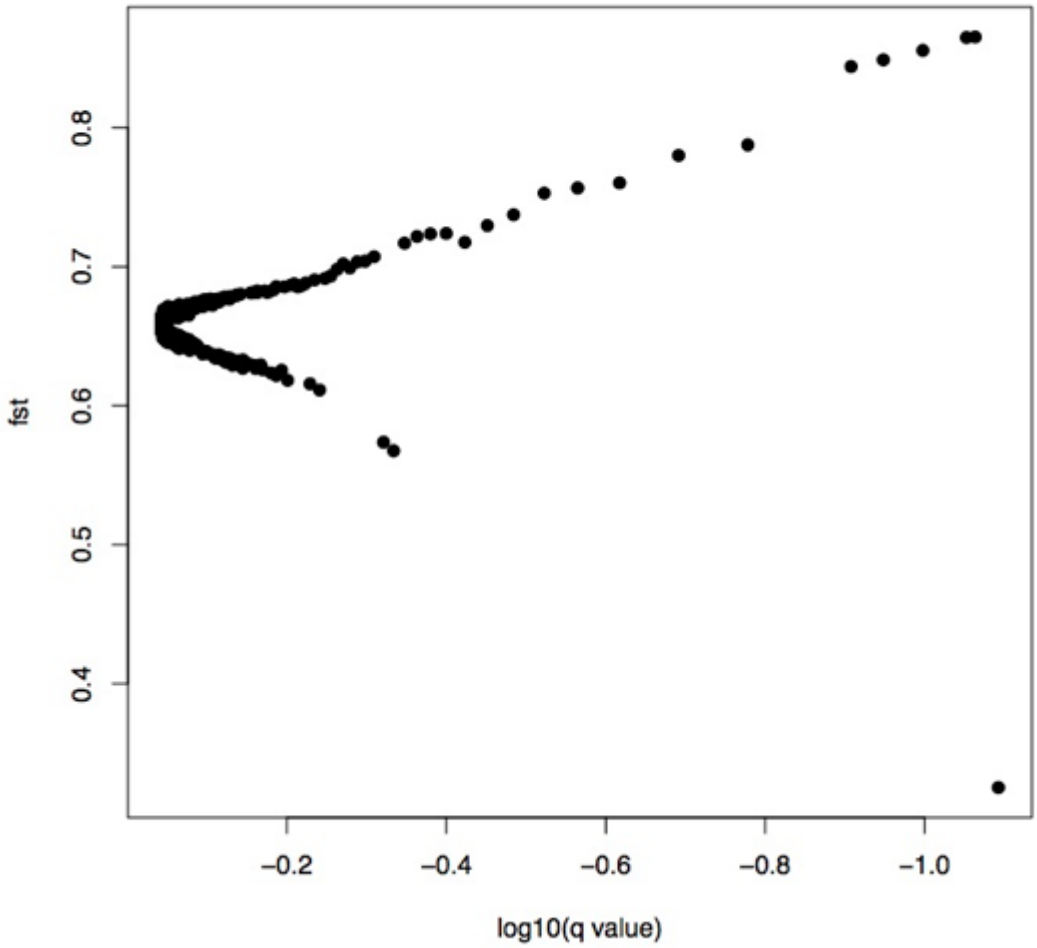
35	CATCCG	P1_35.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATCCGTGCA	P1_35.2	CGGATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
36	CCGTCA	P1_36.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGTCATGCA	P1_36.2	TGACGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
37	GTACGT	P1_37.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACGTTGCA	P1_37.2	ACGTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
38	TAGGCT	P1_38.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGCTTGCA	P1_38.2	AGCCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
39	GGCTAG	P1_39.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCTAGTGCA	P1_39.2	CTAGCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
40	CATGTA	P1_40.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGTATGCA	P1_40.2	TACATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
41	ATTCGG	P1_41.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTCGGTGCA	P1_41.2	CCGAATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
42	TGACCT	P1_42.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACCTTGCA	P1_42.2	AGGTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
43	GCTACT	P1_43.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTACTTGCA	P1_43.2	AGTAGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
44	TCGGTA	P1_44.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGGTATGCA	P1_44.2	TACCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
45	CTGAGG	P1_45.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGAGGTGCA	P1_45.2	CCTCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
46	GCCTTA	P1_46.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTTATGCA	P1_46.2	TAAGGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
47	CGATGT	P1_47.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATGTTGCA	P1_47.2	ACATCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
48	GATTACA	P1_48.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATTACATGCA	P1_48.2	TGTAATCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
49	GGTAGCA	P1_49.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTAGCATGCA	P1_49.2	TGCTACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
50	GTGACCA	P1_50.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGACCATGCA	P1_50.2	TGGTCACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
51	TTATGCA	P1_51.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTATGCATGCA	P1_51.2	TGCATAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

52	ATTGGCA	P1_52.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGGCATGCA	P1_52.2	TGCCAATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
53	TGGTACA	P1_53.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTACATGCA	P1_53.2	TGTACCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
54	GACCTCA	P1_54.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACCTCATGCA	P1_54.2	TGAGGTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
55	TGTGCCA	P1_55.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTGCCATGCA	P1_55.2	TGGCACAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
56	TAGACCG	P1_56.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGACCGTGCA	P1_56.2	CGGTCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
57	GGATTCA	P1_57.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGATTCATGCA	P1_57.2	TGAATCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
58	GATCCAA	P1_58.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCCAATGCA	P1_58.2	TTGGATCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
59	CTGGACA	P1_59.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGACATGCA	P1_59.2	TGTCCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
60	AGACTCG	P1_60.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGACTCGTGCA	P1_60.2	CGAGTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
61	AATTGCG	P1_61.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAATTGCGTGCA	P1_61.2	CGCAATTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
62	TCCAGGA	P1_62.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCAGGATGCA	P1_62.2	TCCTGGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
63	TCAGCAG	P1_63.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAGCAGTGCA	P1_63.2	CTGCTGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
64	CAGTGCA	P1_64.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGCATGCA	P1_64.2	TGCACTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
65	GTACCGA	P1_65.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACCGATGCA	P1_65.2	TCGGTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
66	TGTAACG	P1_66.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTAACGTGCA	P1_66.2	CGTTACAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
67	TACGATA	P1_67.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGATATGCA	P1_67.2	TATCGTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
68	GTAAGCG	P1_68.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAGCGTGCA	P1_68.2	CGCTTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

69	ATGCAAT	P1_69.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCAATTGCA	P1_69.2	ATTGCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
70	CCGGTAA	P1_70.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGGTAATGCA	P1_70.2	TTACCGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
71	AGCTCCG	P1_71.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCTCCGTGCA	P1_71.2	CGGAGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
72	AACTCG	P1_72.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACTCGTGCA	P1_72.2	CGAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
73	AGAATGCA	P1_73.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGAATGCATGCA	P1_73.2	TGCATTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
74	GAATAGCA	P1_74.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAATAGCATGCA	P1_74.2	TGCTATTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
75	ATGAGACA	P1_75.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGACATGCA	P1_75.2	TGTCTCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
76	TGCCACCA	P1_76.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCCACCATGCA	P1_76.2	TGGTGGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
77	ATAGAGCA	P1_77.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGAGCATGCA	P1_77.2	TGCTCTATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
78	ACTCGCCA	P1_78.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTCGCCATGCA	P1_78.2	TGGCGAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
79	TAGGAACA	P1_79.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGAACATGCA	P1_79.2	TGTTCCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
80	GATACGAA	P1_80.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATACGAATGCA	P1_80.2	TTCGTATCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
81	GCACCTCA	P1_81.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCACCTCATGCA	P1_81.2	TGAGGTGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
82	CACTGCCA	P1_82.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACTGCCATGCA	P1_82.2	TGGCAGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
83	ACGATGAA	P1_83.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGATGAATGCA	P1_83.2	TTCATCGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
84	CGCACACT	P1_84.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCACACTTGCA	P1_84.2	AGTGTGCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
85	GGTCTT	P1_85.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTCTTTGCA	P1_85.2	AAGACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

86	CAAGTAGA	P1_86.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAGTAGATGCA	P1_86.2	TCTACTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
87	GCAAGAAT	P1_87.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAAGAATTGCA	P1_87.2	ATTCTTGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
88	ACCTACCG	P1_88.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCTACCGTGCA	P1_88.2	CGGTAGGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
89	CTACCACG	P1_89.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTACCACGTGCA	P1_89.2	CGTGGTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
90	TAGAACGA	P1_90.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGAACGATGCA	P1_90.2	TCGTTCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
91	AGCAGTAA	P1_91.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAGTAATGCA	P1_91.2	TTACTGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
92	GAACTGAA	P1_92.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAACTGAATGCA	P1_92.2	TTCAGTTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
93	ACTCCACG	P1_93.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTCCACGTGCA	P1_93.2	CGTGGAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
94	GAAGACAT	P1_94.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAGACATTGCA	P1_94.2	ATGTCTTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
95	CGGTATGT	P1_95.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGTATGTTGCA	P1_95.2	ACATACCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
96	TCCGCACA	P1_96.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCGCACATGCA	P1_96.2	TGTGCGGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

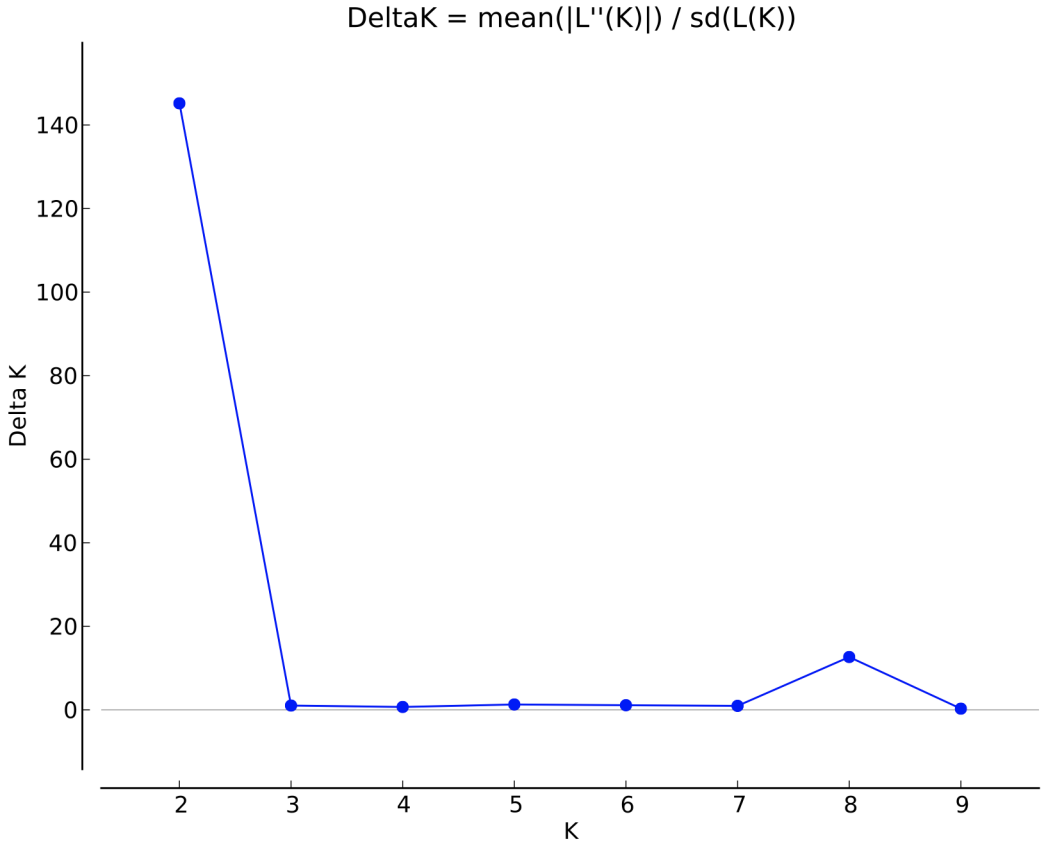
S3: BayeScan 2.1 plot of the global genome scan for all
polymorphic loci. FST is plotted against log10 of the posterior odds, which identifies no outlier markers (candidates for being under positive selection).



S4: Delta K of complete dataset (output from STRUCTURE harvester)

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-128.34	0.7763	NA	NA	NA
2	10	-2.93	0.8718	125.41	126.58	145.186605
3	10	-4.1	0.83	-1.17	0.86	1.036153
4	10	-4.41	0.7923	-0.31	0.55	0.694222

5	10	-4.17	0.7602	0.24	0.98	1.289151
6	10	-4.91	1.3068	-0.74	1.46	1.117252
7	10	-4.19	0.9457	0.72	0.91	0.962258
8	10	-4.38	0.9461	-0.19	11.97	12.651898
9	10	-16.54	38.3964	-12.16	11.09	0.288829
10	10	-17.61	39.9905	-1.07	NA	NA



Chapter 2

Stream flow regime does not predict evolutionary trajectories of diving beetles across complex tropical landscapes

Abstract
Recent theoretical advances have hypothesized a central role of habitat persistence on population genetic structure and resulting biodiversity of freshwater aquatic organisms. Here, I test the hypothesis that lotic (moving water) lineages adapted to comparably geologically stable running water habitats have high levels of endemicity and phylogeographic structure due to the persistent nature of their habitat. I use a nextRAD sequencing approach to investigate the population structure and phylogeography of a widespread New Guinean species of diving beetles, *Philaccolilus ameliae*. The results show that some populations do indeed exhibit fine-scale genetic structure consistent with theoretical predictions associates with stable habitats, other populations show much more complex patterns, including unexpected connectivity across the rugged mountains of New Guinea. These results offer mixed support for the habitat template concept, by which spatial and temporal habitat characteristics drive patterns of life history and dispersal. However, this theory does not fully capture the complexity of the interaction between habitat preference and evolutionary trajectories found in this system. I suggest that predictions of life history variation resulting from dichotomy between lentic and lotic organisms be revised to include more meaningful habitat characterization. Our results also underpin the necessity to study fine-scale processes, as compared to solely macro-ecological patterns, to understand ecological drivers of biodiversity.

BACKGROUND

Understanding how environmental and geological factors shape the geographical distribution of biodiversity is a central theme in biology (Gaston 2000). Despite its importance, many fundamental questions remain, including the degree to which the temporal and spatial nature of available habitats may impact the evolutionary trajectories of species or radiations. The Habitat Template Concept (HTC) has been proposed to explain how the intrinsic features and constraints imposed by a given habitat may drive the evolution of ecological traits and evolutionary strategies of its inhabitants. These drivers, in turn, have scaled effects on the population structure and biogeography of species (Southwood 1977, 1988; Korfiatis & Stamou 1999; see Dijkstra *et al.* 2014 for a review). As a result, the habitat template concept predicts that species that occupy different habitats will exhibit different patterns of ecological and geographical diversification.

According to the HTC, one of the fundamental factors influencing the selection for dispersal is the persistence of habitats through time (Roff 1994; Bohonak 1999). Freshwater ecosystems, as spatially-defined habitats with varied degrees of stability over evolutionary time, are particularly well suited to testing the HTC. *First*, freshwater habitats can be readily categorized, allowing for direct evolutionary comparisons. The major abiotic factor that classifies freshwater habitats is their flow regime: lentic habitats (i.e. lakes, ponds) are those with standing water, and lotic habitats (i.e. streams and rivers) are those with running water (Ribera 2008). *Second*, these two types of habitats further differ in their ecological and spatiotemporal characteristics: Lotic habitats are generally considered to be more stable than lentic habitats as they are more continuous both spatially and temporally (Ribera *et al.* 2001; Dijkstra *et al.* 2014). *Third*, characters under direct selection as a result of habitat associations can be predicted *a priori* for lotic and lentic species: Lotic species are under decreased selection pressure for dispersal owing to the stability of lotic habitats and thus experience a lower risk for local extinction (Roff 1986), than closely related species that are restricted to ephemeral standing water bodies (Dobson & Frid 1998). *Fourth*, because reduced dispersal ability will result in reduced gene flow among populations, increased genetic differentiation of neighboring habitat patches should be observed (Slatkin 1985), a prediction that can be empirically tested (Ribera *et al.* 2001, 2003; Ribera & Vogler 2004). *Finally*, because reduced gene flow leads to an increased probability of peripatric and allopatric speciation, lineages with small geographic range sizes as assumed for lotic species, will have a greater probability of extinction leading to higher species turnover (Ribera *et al.* 2001).

To understand the influence of habitat templates on the evolutionary trajectories of lineages, an association between habitat type and characters under selection must be established. For example, Arribas *et al*. (2012) empirically demonstrated that lentic water beetles have larger wings when compared to closely related lotic species. Reduced vagility leads to a comparatively smaller range size in lotic species (Ribera *et al.* 2001, 2003; Ribera & Vogler 2004; Grewe *et al.* 2013). Ribera and Vogler (2000) demonstrated that among 490 beetle species in the Iberian peninsula, only those with narrow distributions were lotic. The proportion of species diversity was further demonstrated to vary latitudinally, with the proportion of lentic species increasing northward (Ribera *et al.* 2003). This pattern is hypothesized to be the result of a higher dispersal rate among lentic species, leading to faster recolonization of areas glaciated during the Pleistocene.

The genetic structure of freshwater aquatic invertebrate populations is generally stronger in lotic species when compared to their lentic counterparts (i.e. Ribera *et al.* 2001, 2003; Ribera & Vogler 2004, Marten *et al*. (2006), Hof *et al.* 2006, Monaghan *et al.*, 2005)). For example, Hjalmarsson *et al*. (2014) found that tropical riparian (stream margin

inhabiting) beetle species had significantly greater population structure when compared to lentic and lotolentic (generalist) species. Papadopoulou *et al*. (2008, 2009) showed that migration rates influenced variation in the coalescence of mitochondrial DNA (mtDNA) in taxa occupying habitat types that differ in their stability. Using one aquatic and one terrestrial genus of beetles, they suggested that lineages in more stable habitats have greater levels of population sub-division and geographical structure. However, predictions from the HTC have also been challenged by recent molecular genetic studies. For example, a comprehensive phylogeny of Odonata suggested that lentic clades have higher diversification rates than lotic ones, as larger range size also means higher likelihood of vicariant events dividing an ancestral range and creating a higher number of available habitats (Letsch *et al.* 2016). On a larger spatial and temporal scale, Short & Caterino (2009) found different phylogeographic patterns in three sympatric lotic beetle species from three different families, questioning the validity of habitat as a general predictor of evolutionary patterns. However, of the beetles they studied, only one is strictly lotic (water penny *Eubrianax*) while one is found in stream pools (*Stictotarsus*) and another in debris at stream margins (*Anacaena*). Fine scale empirical studies, especially in understudied tropical ecosystems, are needed to investigate to what extent the HTC predicts evolutionary patterns and processes. As the observations above illustrate, it critical to investigate multiple, independent study systems with a focus on habitats of lineages dwelling along water courses. In particular, a large proportion of tropical stream beetle diversity is found in stagnant water at the edge of streams or around the spring (Balke *et al.* 2004).

In this study, I revisit the HTC using the New Guinean endemic genus *Philaccolilus* as a focal group. Species of *Philaccolilus* are strictly running water inhabitants (i.e. lotic), occupying smaller forest creeks, fast flowing montane streams with heavy flooding and streaming, mud free edges of lowland rivers (Balke *et al.* 2000; Fig. 1). The genus comprises 12 described species endemic to New Guinea mainland (Nilsson 2013), and five additional undescribed species from the island that have recently been discovered (Balke, unpublished). Species in the genus have very different range sizes. Most species have small ranges (e.g. endemic to the Papuan Peninsula / isolated mountain ranges), where other morphologically defined species are known to have wider ranges. Here, I focus on one of these widespread species, *Philaccolilus ameliae* (Balke), whose geographical distribution ranges across almost the entire island of New Guinea (Fig. 1). This species includes two subspecies: *P. a. ameliae* (Balke) found from eastern Papua New Guinea to the Bird's Head Peninsula of West Papua, and the morphologically similar *P. a. weylandensis* (Balke) from the Weyland Mountains of Papua. The island of New Guinea has a complex geotectonic history and is composed of numerous geological elements such as continental fragments, oceanic island arcs and a massive central orogen (Toussaint *et al.* 2014), which are expected to have a further impact on species ranges and population connectivity (e.g. Balke *et al.* 2009; Deiner *et al.* 2011; Toussaint *et al.* 2013, 2014), in addition to the ecological preferences of species.

Using a nextRAD (nextera-tagmented, Reductively Amplified DNA) sequencing approach, I investigate the genetic structure of *P. ameliae* across New Guinean rugged landscapes to gain new insights into understanding how habitat characteristics predict patterns of life history and diversification. I specifically test the central prediction of the HTC, that populations of the lotic *P. ameliae* should exhibit low levels of dispersal associated with a high degree of population genetic structure across New Guinean sampled localities.

MATERIAL AND METHODS

Taxon Sampling

I sampled 60 individuals of *Philaccolilus ameliae* from seven localities across New Guinea (Fig. 1), representing both described subspecies (Balke *et al.* 2000). In addition, 28 individuals representing 11 other species of the genus *Philaccolilus* were included to confirm monophyly of *P. ameliae* and to provide a preliminary phylogenetic hypothesis for the genus. In total, the sampling includes 10 of the 12 described species of *Philaccolilus*, in addition to a putative new species (Table S1). One individual from the genus *Laccophilus*, was included as an outgroup. DNA was extracted from whole beetles with punctured metacoxa, using the DNeasy Blood & Tissue Kit from Qiagen (Hilden, Germany). Voucher specimens are housed at the Museum Zoologicum Bogoriense (Cibinong, West Java, Indonesia) and the Zoological State Collection, Munich.

nextRAD Sequencing

I performed population genomic analyses using nextRAD genotyping to collect single nucleotide polymorphism (SNP) data. Using DNA samples, SNPsaurus (SNPsaurus.com) generated nextRAD libraries and performed Illumina sequencing. Genomic DNA was converted into nextRAD libraries as described by Russello et al. (2015). Genomic DNA was first fragmented with Nextera reagent (Illumina, San Diego, USA), which also ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled for fragmenting 10 nanograms of genomic DNA. Fragmented DNA was then PCR amplified, with one of the primers matching the adapter, extending nine nucleotides into the genomic DNA with the selective sequence GTGTAGAGC. Therefore, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer were efficiently amplified by PCR. The nextRAD libraries were sequenced on an Illumina HiSeq 2500 (University of Oregon, USA).

The genotyping analysis used custom scripts (SNPsaurus.com) that created a *de novo* reference from abundant reads, and then mapped all the reads to the reference with an alignment identity threshold of 93% (BBMAP, http://sourceforge.net/projects/bbmap/). Genotype calling was performed using SAMTOOLS and BCFTOOLS (samtools mpileup -gu -Q 10 -t DP,DPR -f ref.fasta -b samples.txt | bcftools call -cv - > genotypes.vcf). The vcf files were converted to PHYLIP format by concatenating the *de novo* reference and substituting the called genotypes for each sample at the polymorphic positions.

I tested all polymorphic loci for signals of selection using the Bayesian simulation method of Beaumont and Balding (2004) as implemented in BAYESCAN 2.1. (Foll & Gaggiotti 2008). Analyses were run separately for each cluster studied in detail (described below). I used a prior odds value of 10, with 100,000 iterations and a burn-in of 50,000 iterations. I identified loci that were significant outliers at a *q*-value of 0.20. Since most population genetics algorisms assumes neutral markers, loci identified as statistical outliers were removed from the dataset before further analyses.

Sequencing of the nextRAD library produced a total of 130,005,273 reads from 87 individuals, and the reads collapsed to 40,059 loci that were distinct from other loci by an identity threshold of at least 92%. These loci were used as a *de novo* reference for aligning the sequence reads from each sample using BBMAP (k=9, slow mode, indel=15, minid=.92). The resulting bam files were converted to a vcf genotype table using SAMTOOLS mpileup. The putative variants in the vcf genotype table were then filtered using vcftools to remove variants not present in at least 85%, 95% of the samples and allowing variants with a population frequency of at least 0.05 to reduce artefactual variants. After filtering, the final dataset consisted of 5,609 SNPs in 1,726 loci across 88 individuals. The vcf file was

converted to PHYLIP format by concatenating the full sequence of each locus and then for each sample substituting the alleles found to create a sample-specific sequence. Thus, it contained the variant and invariant nucleotides for each sample. From the independent SNPs tested for statistical outlier, BAYESCAN determined that no loci displayed signal of selection in any of the three clusters, therefore all 5,609 SNPs were retained for population genetic analyses.

Phylogenetic analyses

I conducted phylogenetic analyses by generating consensus sequences of all individuals of each species in SEQUENCHER 5.0.1 (GeneCodes Corp., Ann Arbor, MI, USA). The best partitioning scheme was selected in PARTITIONFINDER 2 (Lanfear *et al.* 2016) using the 'greedy' algorithm. In the gene tree analyses, character sets were defined according to the number of phylogenetically informative sites, giving a total of 10 character sets. The maximum likelihood (ML) analyses were performed using RAxML 8.0.19 (Stamatakis 2014) with a GTR+G model for each partition, conducting 20 independent tree searches for the best ML tree. I assessed support for the best ML topology by performing nonparametric bootstrapping using the autoMRE option in RAxML.

I also performed partitioned concatenated Bayesian inference analyses with MRBAYES 3.2 (Ronquist *et al.* 2012). The best partitioning scheme and corresponding models of substitution were determined using PartitionFinder 2. The analysis consisted of two independent runs of four Markov chains Monte Carlo (MCMC) running 20 million generations with a tree and parameter sampled every 1000 generations. After removing 30% of the posterior samples as burn-in, I generated a 50% majority-rule consensus topology. The posterior values were examined in TRACER v1.0.6 (Rambaut *et al.* 2014) to ensure that the Effective Sample Size (ESS) of each parameter was sufficiently sampled.

Population genetics analyses

I assessed overall population structuring using a Bayesian clustering analysis in STRUCTURE 2.3.4 (Pritchard *et al.* 2000). Run length was set to 100,000 MCMC replicates after a burn-in period of 1,000,000 using correlated allele frequencies under a straight admixture model. The number of clusters (K) was varied from 1 to 8, with 10 replicates for each value of K. I used the admixture model with correlated allelic frequencies. The broad scale number of clusters was initially determined by examining both the posterior probabilities of the data for each K and the ΔK estimator described by Evanno et al. (2005) as calculated in Structure Harvester (Earl & vonHoldt 2012). Results for the identified optimal values of K were summarized using CLUMPP ver. 1.1 (Jakobsson & Rosenberg 2007) using 1000 permutations and the LargeKGreedy algorithm. The result was then plotted using DISTRUCT ver. 1.1 (Rosenberg 2004).

Initial results suggested divergent clusters within *P. ameliae* (described in detail in the Results section below). To detect population subdivision that may be overlooked within each cluster (potential cryptic species), I subsequently conducted STRUCTURE analyses for each subclade, as in Gowen *et al.* (2014).

In addition, structuring was examined using principal component analyses (PCA), a multivariate approach. Specifically, I used the Excel based program GENALEX 6.1 (Peakall & Smouse 2006) to calculate a genetic distance and to convert this into a covariance matrix with data standardization for the PCA. The first three principal components were plotted in the R package scatterplot3d (Ligges & Martin 2003). Each cluster was analyzed separately. Levels of genetic differentiation among groups of each of the three clusters (see Results) were estimated by pairwise F_{ST} (Weir & Cockerham 1984) as implemented in GENETIX (Belkhir *et al.* 2004) using 1,000 permutations.

To determine the extent to which genetic variation was partitioned across samples within each clade, I conducted a hierarchical analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) using ARLEQUIN 3.5 (Excoffier & Lischer 2010) with significance assessed using 1023 permutations.

RESULTS

Molecular phylogenetics

The MRBAYES analyses converged well as indicated by an average standard deviation of split frequency of 0.000758 after removing burn-in samples. All parameters had ESS values above 200. The results of MRBAYES and RAXML analyses are presented in Figure 2. I recover populations of *P. ameliae* within three separate clades. All analyses recover *P. ameliae* as paraphyletic with respect to *P. mas*, which is recovered as sister to clade (A). In this paper, I will refer to the clade comprising all *P. ameliae*-like-beetles (including *P. mas*) as the "*P. ameliae* complex". Taxonomic issues will be treated in a separate taxonomic study. The three main *P. ameliae* monophyletic clades consist of individuals from (A) the Bird's Head peninsula (Arfak, Kebar, Testega), (B) Foja mountains and Sandaun province, and (C) Ok Sibil (Fig. 2).

Population genetics

Initial STRUCTURE analyses including all individuals of the *P. ameliae* complex recovered the same clusters described above (Fig. 3). Since these clusters seem to act as independent evolutionary units (putative cryptic species), each cluster was analysed separately in subsequent population genetic analyses. The single *P. mas* individual makes up a unique lineage that is clearly distinct from all other samples included, therefore, it was not included in further population genetic analyses. Analyzing each group separately, the ΔK method identified two genetic populations in each group (Fig. 3). However, using the method of Evanno et al. (2005) for estimating the optimal number of cluster(s) based on ΔK, it would not be possible to select a scenario of K=1. Moreover, Latch et al. (2006) showed that STRUCTURE could not distinguish clusters and delivered inconsistent results when populations have an F_{ST} of 0.03 or below. Therefore, I distinguish between scenarios of two discrete populations and one panmictic population by observing the level of admixture.

Both clades (A) and (B) have very weak structure and unclear assignments of individuals into clusters, suggesting that each group forms a single population (Fig. 3). In the Bird's Head group (A), the single individual from Arfak is distinct from individuals of the Kebar populations, but there is high levels of mixed assignment between the Kebar and Testega populations. In the Foja + Sandaun group (B), there is high level of admixture between the Foja and Sandaun populations. In group (C), the Ok Sibil and Weyland mountains populations remain distinct with no observable admixture.

The AMOVA revealed a low amount of variation in group A (10%, $p = 0.1$) (Table 1). Pairwise F_{ST} estimations showed congruent results with AMOVA and STRUCTURE, suggesting insignificant differentiation between the three populations in the Bird's Head peninsula (0.02-0.03). The PCA analysis showed that the first three components explained only a minor portion of the total variance (20.82% cumulative). Overall, our PCA analyses are congruent in showing no distinct clustering between the Kebar and Testega populations (Fig. 4). Similarly, the AMOVA and pairwise F_{ST} analyses revealed a limited amount of variation between the two populations of *P. ameliae* found in Foja and Sandaun (variation among group=27%, $p = 0.045$; $F_{ST} = 0.03$) (Tables 1). PCA showed a similar pattern with high level of admixture, the first three principal components explaining 44% of the total variation (Fig. 4). In contrast, the AMOVA and pairwise F_{ST} analyses confirmed the high

level of variation between the two populations of *P. ameliae* found in Ok Sibil and the Weyland mountains (variation among group=52%, $p = 0.00587$; $F_{ST} = 0.34$) (Tables 1&2). PCA also showed two distinct clusters (Fig. 4).

DISCUSSION

New Guinea is a geologically complex, tropical environment, with a large central mountain range that divides the island in the middle, as well as numerous mountain ranges (e.g. along the north coast), with uncounted watersheds and stream systems. Despite the unique biota and paleo-geological history of the island, evolutionary studies are rare, principally due to the difficulty to collect samples (but see e.g. Deiner *et al.* 2011; Georges *et al.* 2014; Unmack *et al.* 2013; Toussaint *et al.* 2014). This lack of information is especially true at the intraspecific level, with only a handful of phylogeographic works in the New Guinean region (e.g. Toussaint *et al.* 2013; Janda *et al.* 2016). As a result, the micro- and macroevolutionary processes leading to the astonishing biodiversity of the island heretofore are mostly unknown. Following the HTC, one can predict that New Guinean lotic lineages should exhibit extreme levels of local endemism and population structure between watersheds and across geological terranes on the island. Such a pattern has been previously suggested by several studies involving large time-calibrated molecular phylogenies, including riparian *Exocelina* diving beetles (Toussaint *et al.* 2014, 2015), rainbowfishes (Unmack *et al.* 2013) and the New Guinea snapping turtle *Elseya novaeguineae* (Georges *et al.* 2014). However, there are no studies to date that focus on fine-scale phylogeographic patterns among populations of aquatic lineages across the island (but see Georges *et al*. 2014 for Miocenic divergence of regional populations).

Morphological systematic work has categorized the three clades of the *P. ameliae* complex as a single widespread species ranging across the entire island (Fig. 1). However, the genomic nexRAD data used in this study reveal strong geographical subdivision consistent with predictions of the HTC. Yet, within two of the three genetic clusters, I find astonishing levels of connectivity across complex landscapes and at different elevations (e.g., Foja localities are 150m high whereas Sandaun localities are at 700m high). The Bird's Head cluster (A) includes samples from different watersheds on different geological elements (i.e., Arfak and Tamrau), yet our data reveal a high level of genetic exchange across the entire peninsula (Fig. 3). Populations of cluster (B) also originate from different catchments and are even more geographically and geologically distinct from each other, yet the same pattern of high connetivity is found (Fig. 3). The Sandaun localities in Papua New Guinea are situated around the central orogeny while the Foja locality in West Papua is at the northern foot of a different geological element, known as the oceanic Gauttier terrane (see Toussaint *et al.* 2014). Thus, the results suggest genetic exchange between highly disjunct localities.

In contrast, I find no evidence of genetic exchange between the two populations of Ok Sibil and Sandaun although these two localities are only about 120 kilometers apart and are both located in the central orogeny (different watershed, Sandaun north and Ok Sibil south of the central divide) (Fig. 3). Importantly, samples from Ok Sibil and the Weyland mountains far west, and north of the central divide make up a separate cluster in all the analyses, distinct from other *P. ameliae* complex populations.

lineage idiosyncratic patterns within the *P. ameliae* species complex was found; from astonishing high level of connectivity across a broad geographical range, to absence of genetic exchange between neighboring populations. These differences might be due in part to a range of temporal processes underlying population divergence, in which some populations are already isolated, while others still show signatures of recent dispersal across major landforms (e.g., Foja / Sandaun).

Conclusion

Using a genomic approach to investigate the phylogeography of a supposedly widespread species of lotic beetle, putatively cryptic species and complex patterns of population structure was discovered. Some clusters of *P. ameliae* form panmictic populations throughout a large geographic range and others lacking genetic exchange. Therefore, some populations reflect the predictions of the HTC while others do not. This suggests that the HTC does not capture the complexity of ecological preferences and their relation to population structuring processes in nature. Our study indicates that water flow alone is insufficient for predicting genetic structure and level of dispersal in aquatic systems. It seems that the lentic/lotic dichotomy might be an overly simplistic system and that other factors such as ecological microhabitat affinity, geographical factors, and organisms' life history traits may play a significant role in the evolutions of the species' propensity to disperse (Short & Caterino 2009). The high connectivity recovered in our study might also reflect the state of very recent range expansion (e.g. Toussaint et al. 2013). The temporal dimension is crucial to gain new insights into the evolutionary history of population and metapopulation structuring across geographical landscapes. Further efforts should focus on combining fine-scale ecological data with comprehensive genomic datasets in a time-calibrated framework. Adding to our knowledge on the interface between ecology and evolution in aquatic biology will likely require such an integrative approach.

REFERENCES

Arribas P, Velasco J, Abellán P *et al.* (2012) Dispersal ability rather than ecological tolerance drives differences in range size between lentic and lotic water beetles (Coleoptera: Hydrophilidae). *Journal of Biogeography*, **39**, 984–994.

Baird N, Etter P, Atwood T, Currey M (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, **3**, e3376.

Balke M, Jäch MA, Hendrich L (2004) Coleoptera. In: *Freshwater invertebrates of the Malaysian region* (eds Yule CM, Yong HS), pp. 555–609. Akademi Sains Malaysia, Kuala Lumpur, Malaysia.

Balke M, Larson D, Hendrich L, Konyorah E (2000) A revision of the New Guinea water beetle genus Philaccolilus Guignot, stat.n. (Coleoptera, Dytiscidae). *Deutsche Entomologische Zeitschrift*, **47**, 29–50.

Balke M, Ribera I, Hendrich L *et al.* (2009) New Guinea highland origin of a widespread arthropod supertramp. *Proceedings of the Royal Society of London B: Biological Sciences*.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, Population genetics software for Windows TM.

Bilton DT, Freeland JR, Okamura B (2001) Dispersal in Freshwater Invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.

Bohonak A (1999) Dispersal, gene flow, and population structure. *Quarterly review of biology*, 21–45.

Bowler DE, Benton TG (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological reviews of the Cambridge Philosophical Society*, **80**, 205–25.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.

Chifman J, Kubatko L (2015) Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, **374**, 35–47.

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–340.

Deiner K, Lemmon AR, Mack AL, Fleischer RC, Dumbacher JP (2011) A Passerine Bird's Evolution Corroborates the Geologic History of the Island of New Guinea (B Schierwater, Ed,). *PLoS ONE*, **6**, e19479.

Denno R, Olmstead K, McCLoud E (1989) Reproductive cost of flight capability: a comparison of life history traits in wing dimorphic planthoppers. *Ecological Entomology*, **14**, 31–44.

Dijkstra K-DB, Monaghan MT, Pauls SU (2014) Freshwater Biodiversity and Aquatic Insect Diversification. *Annual Review of Entomology*, **59**, 143–163.

Dobson M, Frid C (1998) Ecology of aquatic ecosystems. *Essex: Addison Wesley Longman Limited*.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for

visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.
Edwards JS, Sugg P (1993) Arthropod fallout as a resource in the recolonization of Mount St. Helens. *Ecology*, **74**, 954–958.
Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, **107**, 16196–16200.
Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic biology*, **52**, 665–673.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplo- types: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.
Gaston KJ (2000) Global patterns in biodiversity. *Nature*, **405**, 220–227.
Georges A, Spencer R, Welsh M *et al.* (2011) Application of the precautionary principle to taxa of uncertain status: the case of the Bellinger River turtle. *Endangered Species Research*, **14**, 127–134.
Gottscho AD, Wood DA, Vandergast AG *et al.* (2017) Lineage diversification of fringe-toed lizards (Phrynosomatidae: Uma notata complex) in the Colorado Desert: Delimiting species in the presence of gene flow. *Molecular Phylogenetics and Evolution*, **106**, 103–117.
Gowen FC, Maley JM, Cicero C *et al.* (2014) Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC evolutionary biology*, **14**, 135.
Grewe Y, Hof C, Dehling DM, Brandl R, Brändle M (2013) Recent range shifts of European dragonflies provide support for an inverse relationship between habitat predictability and dispersal. *Global Ecology and Biogeography*, **22**, 403–409.
Gronau I, Hubisz M, Gulko B, Danko C, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics*.
Hey J, Pinho C (2012) Population genetics and objectivity in species diagnosis. *Evolution*, **66**, 1413–1429.
Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**, 182–192.
Hipp AL, Eaton DAR, Cavender-Bares J *et al.* (2014) A framework phylogeny of the American oak clade based on sequenced RAD data (S Joly, Ed,). *PLoS ONE*, **9**, e93975.
Hjalmarsson A, Bergsten J, Monaghan M (2014) Dispersal is linked to habitat use in 59 species of water beetles (Coleoptera: Adephaga) on Madagascar. *Ecography*, **38.7**, 732–739.
Hof C, Brändle M, Dehling DM *et al.* (2012) Habitat stability affects dispersal and the ability to track climate change. *Biology Letters*, **8**.

Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian Phylogenetic Model Selection Using Reversible Jump Markov Chain Monte Carlo. *Molecular Biology and Evolution*, **21**, 1123–1133.

Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Janda M, Matos-Marav, P, Borovanska M *et al.* (2016) Phylogeny and population genetic structure of the ant genus Acropyga (Hymenoptera : Formicidae) in Papua New Guinea. *Invertebrate Systematics*, **30**, 28.

Jones J, Fan S, Franchini P (2013) The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular ...*, **22**, 2986–3001.

Knowles LL, Carstens B (2007) Delimiting Species without Monophyletic Gene Trees. *Systematic Biology*, **56**, 887–895.

Korfiatis KJ, Stamou GP (1999) Habitat Templets and the Changing Worldview of Ecology. *Biology and Philosophy*, **14**, 375–393.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution*, msw260.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295–302.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society of London B: Biological Sciences*.

Leache AD, Fujita MK, Minin VN, Bouckaert RR (2014a) Species delimitation using genome-wide SNP Data. *Systematic Biology*, **63**, 534–542.

Leache AD, Harris RB, Rannala B, Yang Z (2014b) The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Systematic Biology*, **63**, 17–30.

Letsch H, Gottsberger B, Ware JL (2016) Not going with the flow: a comprehensive time-calibrated phylogeny of dragonflies (Anisoptera: Odonata: Insecta) provides evidence for the role of lentic habitats on diversification. *Molecular Ecology*, **25**, 1340–1353.

Ligges U, Martin M (2003) Scatterplot3d – an R package for Visualizing Multivariate Data. *Cran*, **8**, 1–36.

Lim GS, Balke M, Meier R (2012) Determining Species Boundaries in a World Full of Rarity: Singletons, Species Delimitation Methods. *Systematic Biology*, **61**, 165–169.

Maddison WP (1997) Gene trees in species trees. *Systematic biology*, **46**, 523–536.

Maddison WP, Knowles LL (2006) Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology*, **55**, 21–30.

Maddison WP, Maddison DR (2008) Mesquite: A modular system for evolutionary analysis. *Evolution*, **62**, 1103–1118.

Marten A, Brandle M, Brandle R (2006) Habitat type predicts genetic population differentiation in freshwater invertebrates. *Molecular Ecology*, **15**, 2643–2651.

Minh BQ, Nguyen MAT, Von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, **30**, 1188–1195.

Mirarab S, Reaz R, Bayzid MS *et al.* (2014) ASTRAL: Genome-scale coalescent-based species tree estimation. In: *Bioinformatics*, pp. i541–i548. Oxford University Press.
Monaghan MT, Gattolliat J-L, Sartori M *et al.* (2005) Trans-oceanic and endemic origins of the small minnow mayflies (Ephemeroptera, Baetidae) of Madagascar. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**.
Myers EA, Rodríguez-Robles JA, Denardo DF *et al.* (2013) Multilocus phylogeographic assessment of the California Mountain Kingsnake (Lampropeltis zonata) suggests alternative patterns of diversification for the California Floristic Province. *Molecular Ecology*, **22**, 5418–5429.
Nadeau NJ, Martin SH, Kozak KM *et al.* (2013) Genome-wide patterns of divergence and gene flow across a butterfly radiation. In: *Molecular Ecology*, pp. 814–826.
Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.
Nilsson AN (2013) *A World Catalogue of the Family Dytiscidae , or the Diving Beetles (Coleoptera , Adephaga)*.
Papadopoulou A, Anastasiou I (2009) Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: the effect of dispersal ability and habitat preference. *Molecular Ecology*, **18**, 2503–2517.
Papadopoulou A, Bergsten J, Fujisawa T *et al.* (2008) Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 2987–2996.
Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Pyron RA, Hsieh FW, Lemmon AR, Lemmon EM, Hendry CR (2016a) Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (Storeria). *Zoological Journal of the Linnean Society*, **177**, 937–949.
Pyron RA, Hsieh FW, Lemmon AR, Lemmon EM, Hendry CR (2016b) Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). *Zoological Journal of the Linnean Society*, **177**, 937–949.
De Queiroz K (1999) *The general lineage concept of species and the defining properties of the species*. MIT Press.
De Queiroz K (2005) A Unified Concept of Species and Its Consequences for the Future of Taxonomy. *Proceedings of the California Academy of Sciences. I)*, **4**.
De Queiroz K (2007) Species Concepts and Species Delimitation. *Systematic Biology*, **56**, 879–886.
Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer | BEAST.
Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–56.
Reeves PA, Richards CM (2011) Species Delimitation under the General Lineage Concept: An Empirical Example Using Wild North American Hops (Cannabaceae: Humulus lupulus). *Systematic Biology*, **60**, 45–59.
Ribera I (2008) Habitat constraints and the generation of diversity in freshwater macroinvertebrates. In: *Aquatic insects: challenges to populations* (eds Lancaster J, Briers RA), pp. 289–311. CABI, Wallingford.

Ribera I, Dolédec S, Downie I, Foster G (2001) Effect of land disturbance and stress on species traits of ground beetle assemblages. *Ecology*, **82**, 1112–1129.
Ribera I, Foster G, Vogler A (2003) Does habitat use explain large scale species richness patterns of aquatic beetles in Europe? *Ecography*, **26**, 145–152.
Ribera I, Vogler P (2000) Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera. *Biological Journal of the Linnean Society*, **71**, 33–52.
Ribera I, Vogler A (2004) Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, **13**, 179–193.
Riedel A, Sagata K, Surbakti S, Rene Tänzler, Michael Balke (2013) One hundred and one new species of Trigonopterus weevils from New Guinea. *ZooKeys*, 1–150.
Roff D (1986) The evolution of wing dimorphism in insects. *Evolution*, **40**, 1009–1020.
Roff D (1994) Habitat persistence and the evolution of wing dimorphism in insects. *American Naturalist*, **144**, 772–798.
Ronquist F, Teslenko M, Van Der Mark P *et al.* (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ*, **3**, e1106.
Short AEZ, Caterino MS (2009) On the validity of habitat as a predictor of genetic structure in aquatic systems: A comparative study using California water beetles. *Molecular Ecology*, **18**, 403–414.
Slatkin M (1985) Gene flow in natural populations. *Annual review of ecology and systematics*, 393–430.
Southwood T (1977) Habitat, the templet for ecological strategies? *The Journal of Animal Ecology*, **64**, 337–365.
Southwood TRE (1988) Tactics, strategies and templets. *Oikos*, **52**, 3–18.
Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 1607–1612.
Toussaint EFA, Hall R, Monaghan MT *et al.* (2014) The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature communications*, **5**, 4001.
Toussaint EFA, Hendrich L, Shaverdo H, Balke M (2015) Mosaic patterns of diversification dynamics following the colonization of Melanesian islands. *Scientific Reports*, **5**, 16016.
Toussaint EFA, Sagata K, Surbakti S, Hendrich L, Balke M (2013) Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. *Ecology and Evolution*, **3**, 1031–1049.
Unmack PJ, Allen GR, Johnson JB (2013) Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. *Molecular Phylogenetics and Evolution*, **67**, 15–27.
Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. In: *Molecular Ecology*, pp. 787–798.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.
Wendel J, Doyle J (1998) Phylogenetic incongruence: window into genome history and molecular evolution. *Molecular systematics of plants II*, 265–296.
Will K, Mishler B, Wheeler Q (2005) The Perils of DNA Barcoding and the Need for Integrative Taxonomy. *Systematic Biology*, **54**, 844–851.
Wright S (1931) Evolution in Mendelian populations. *Genetics*.
Yang Z (2015) The BPP program for species tree estimation and species delimitation. *Current Zoology*, **61**, 854–865.
Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264–9269.
Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple Loci. *Molecular biology and evolution*, **31**, 3125–35.
Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, **29**, 2869–2876.
Zhang C, Zhang D-X, Zhu T, Yang Z (2011) Evaluation of a Bayesian Coalescent Method of Species Delimitation. *Systematic Biology*, **60**, 747–761.

Data Accessibility

Raw nextRAD sequences are deposited in Dryad : doi:10.5061/dryad.6gs26

TABLES

Table 1. Analyses of molecular variance (AMOVA) of population pairs within clades A, B, and C of the species complex.

AMOVA of *P. ameliae* Clade A

Source of variation	*d.f.*	*Sum of squares*	*Percentage of variation*	*Fixation index*
Among group	2	453.22	9.98	F_{ST} = 0.10, p=0.00098
Within populations	69	4475.33	90.02	

AMOVA of *P. ameliae* Clade B

Source of variation	*d.f.*	*Sum of squares*	*Percentage of variation*	*Fixation index*
Among group	1	130.81	26.74	F_{ST} =0.27, p= 0.045
Within populations	26	566.33	73.26	

AMOVA of *P. ameliae* Clade C

Source of variation	*d.f.*	*Sum of squares*	*Percentage of variation*	*Fixation index*
Among group	1	241.55	51.66	F_{ST} = 0.51, p= 0.00587
Within populations	32	904.8	48.34	

FIGURE LEGENDS

Figure 1. Distribution of *Philaccolilus ameliae* across New Guinea.

Map of Melanesia with detailed sampling localities. Distribution according to Balke et al. 2000 and Museum collection data. Two pictures of *Philaccolilus ameliae* typical habitat are presented at the bottom of the figure. The base map was generated in the Google Maps API's StylingWizard (https://mapstyle.withgoogle.com/) and edited in a graphic design software.

Figure 2. Phylogenetic relationships as recovered in the MRBAYES and RAXML analyses of the concatenated nextRAD dataset.

Color boxes at each node are bootstrap support value of the RAXML analysis and posterior probability of the MRBAYES analysis. Tip labels indicate species followed by location.

Figure 3. Population genetics of the *Philaccolilus ameliae* species complex.

(i) A map of New Guinea with collecting localities of the eight populations studied, (locality colors correspond to colors in STRUCTURE barplots; (ii) Bayesian clustering analyses of nextRAD SNPs data in STRUCTURE. Barplots from STRUCTURE runs for (iia) the *P. ameliae* species complex (*K=1-10*) and (iib) the genetic clusters A, B and C separately ($K = 2$ for each successive run). Showing that the complex is split into three distinct clusters with no genetic admixture. There is a high level of admixture within clusters A and B. The two populations in cluster C remain discrete. The base map was generated in the Google Maps API's StylingWizard (https://mapstyle.withgoogle.com/) and edited in a graphic design software.

Figure 4. Three-dimensional plot of a Principal Coordinates Analysis based on individual nextRAD genotypes. Individuals are color-coded according to collection locality. Table indicates percentage of variation explained by the first three axes.

FIGURES
Figure 1.

Figure 2.

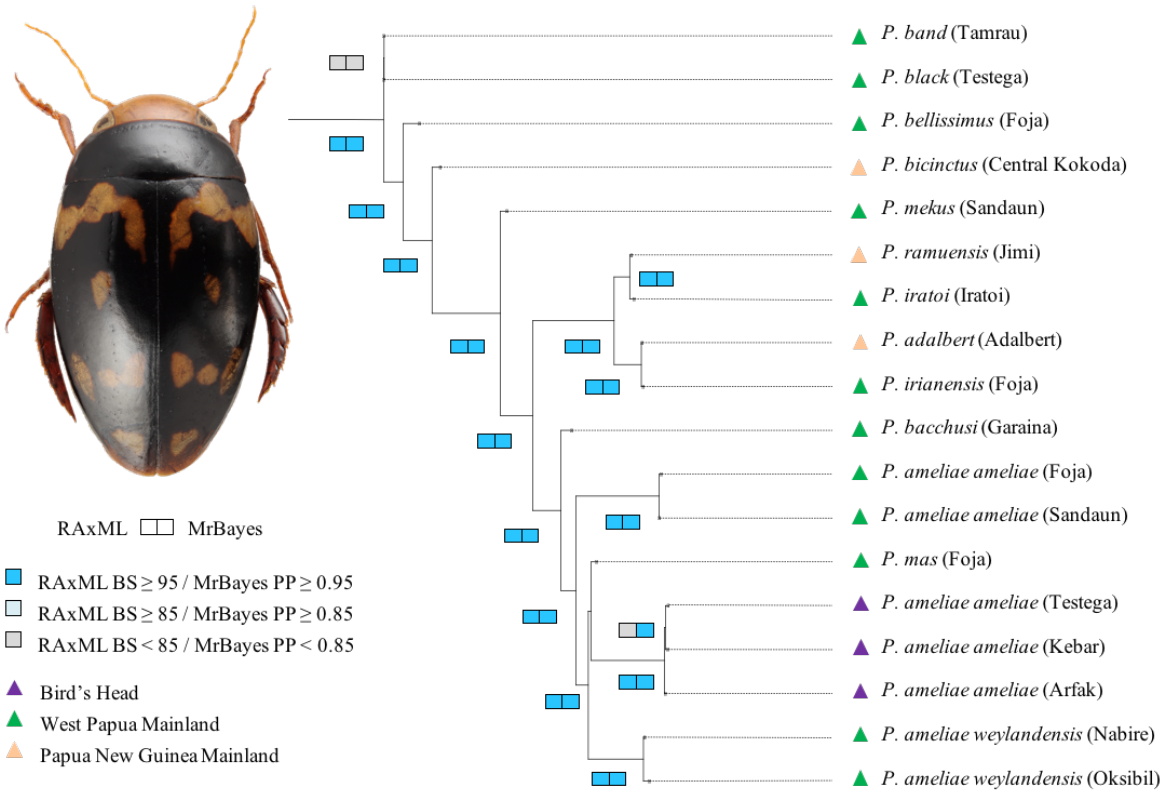

RAxML MrBayes
RAxML BS ≥ 95 / MrBayes PP ≥ 0.95
RAxML BS ≥ 85 / MrBayes PP ≥ 0.85
RAxML BS < 85 / MrBayes PP < 0.85
Bird's Head
West Papua Mainland
Papua New Guinea Mainland
P. band (Tamrau)
P. black (Testega)
P. bellissimus (Foja)
P. bicinctus (Central Kokoda)
P. mekus (Sandaun)
P. ramuensis (Jimi)
P. iratoi (Iratoi)
P. adalbert (Adalbert)
P. irianensis (Foja)
P. bacchusi (Garaina)
P. ameliae ameliae (Foja)
P. ameliae ameliae (Sandaun)
P. mas (Foja)
P. ameliae ameliae (Testega)
P. ameliae ameliae (Kebar)
P. ameliae ameliae (Arfak)
P. ameliae weylandensis (Nabire)
P. ameliae weylandensis (Oksibil)

Figure 3.


Equator
Kebar
Arfak
South Pacific Ocean
Testega
Foja Mts.
Bird's Head
South Nabire
Sandaun
Banda Sea
Weyland Mts.
Ok Sibil
N
W
E
S
Arafura Sea
100 km

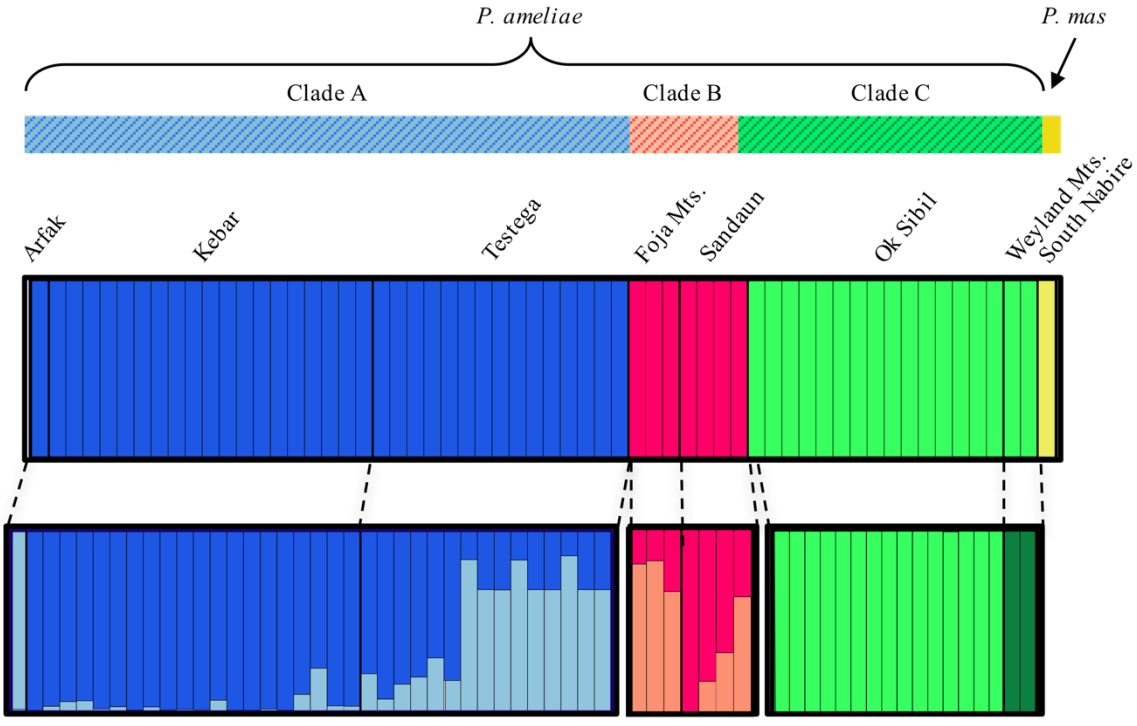
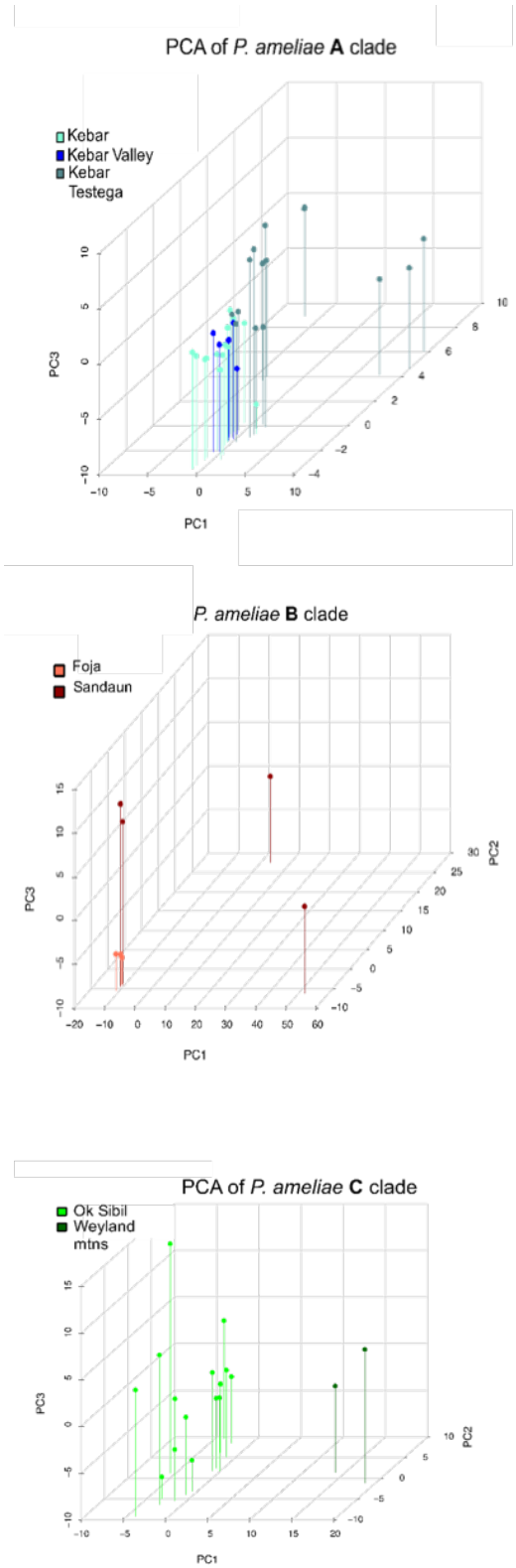

P. ameliae
P. mas
Clade A
Clade B
Clade C
Arfak
Kebar
Testega
Foja Mts.
Sandaun
Ok Sibil
Weyland Mts.
South Nabire

Figure 4.


PCA of *P. ameliae* **A** clade
Kebar
Kebar Valley
Kebar Testega
PC1
PC2
PC3
P. ameliae **B** clade
Foja
Sandaun
PCA of *P. ameliae* **C** clade
Ok Sibil
Weyland mtns

SUPPLEMENTARY MATERIAL

Table S1. Taxon sampling used in this study

Genus	Species	Code	Country	Province	Locality
Philaccolilus	*ameliae*	MB6883	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6884	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6885	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6886	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6887	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6890	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6891	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6892	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6893	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6894	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6895	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6896	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6897	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB3053	Indonesia	West Papua	Arfak
Philaccolilus	*ameliae*	MB6461	Indonesia	Papua	Foja
Philaccolilus	*ameliae*	MB6462	Indonesia	Papua	Foja
Philaccolilus	*ameliae*	MB6851	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6852	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6853	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6854	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6855	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6856	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB3735	PNG	Sandaun	Mianmin
Philaccolilus	*ameliae*	MB3736	PNG	Sandaun	Mianmin
Philaccolilus	*ameliae*	MB4939	PNG	Sandaun	Mianmin area
Philaccolilus	*ameliae*	MB4940	PNG	Sandaun	Mianmin area
Philaccolilus	*ameliae*	MB4207	Indonesia	Papua	Road Nabire-Enarotali KM 95
Philaccolilus	*ameliae*	MB4208	Indonesia	Papua	Road Nabire-Enarotali KM 95
Philaccolilus	*ameliae*	MB6924	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6925	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6926	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6927	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6928	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6929	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6930	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv

Philaccolilus	*ameliae*	MB6931	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6932	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6933	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6934	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6935	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6936	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6937	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6938	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6857	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6858	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6898	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6899	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6900	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6901	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6902	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6903	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6904	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6905	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6906	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6907	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6908	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6909	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6910	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6911	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6912	Indonesia	Papua Barat	Testega
Philaccolilus	*bacchusi*	MB3835	PNG	Morobe	Garaina
Philaccolilus	*bacchusi*	MB3838	PNG	Morobe	Garaina
Philaccolilus	*bellissimus*	MB6464	Indonesia	Papua	Sarmi, Waaf, N Foja Mts, riverbank
Philaccolilus	*bicinctus*	MB4116	PNG	Central	Kokoda Trek
Philaccolilus	*bicintctus*	MB2845	PNG	National Capital District	Varirata NP
Philaccolilus	*black*	MB6227	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream
Philaccolilus	*black*	MB6228	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream
Philaccolilus	*black band*	MB6915	Indonesia	Papua Barat	Testega
Philaccolilus	*Incognitus*	MB4087	PNG	Madang	Highway nr Madang, ford
Philaccolilus	*Iratoi*	MB6916	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6917	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6918	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6919	Indonesia	Papua	S Iratoi, forest

Philaccolilus	*Iratoi*	MB6920	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6921	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6922	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6923	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*irianensis*	MB3046	PNG	Sandaun	Bewani Stn.,base of Bewani Mts
Philaccolilus	*irianensis*	MB5139	PNG	Papua	Road Nabire-Enarotali KM 108
Philaccolilus	*irianensis*	MB6459	Indonesia	Papua	Sarmi, Waaf, N Foja Mts, riverbank
Philaccolilus	*irianensis*	MB6460	Indonesia	Papua	Sarmi, Waaf, N Foja Mts, riverbank
Philaccolilus	*mas*	MB3043	Indonesia	Papua	Road Nabire - Ilaga, KM 117
Philaccolilus	*mekus*	MB4936	PNG	Sandaun	Mianmin area
Philaccolilus	*mekus*	MB4937	PNG	Sandaun	Mianmin area
Philaccolilus	*mekus*	MB6025	PNG	Sandaun	Mianmin area
Philaccolilus	*ramuensis*	MB2871	PNG	Western Highlands	Jimi Valley, Sendiap Station
Philaccolilus	*ramuensis*	MB6400	PNG	Western Highlands	Jimi Valley, Sendiap Station
Philaccolilus	*ramuensis*	MB6401	PNG	Western Highlands	Jimi Valley, Sendiap Station

Note: PNG, Papua New Guinea; KM, kilometer; Mts, mountains; S, South.

Chapter 3

Species delimitation in rare and diverse tropical stream beetles across rugged and remote tropical landscapes.

ABSTRACT
Tropical invertebrates constitute much of the world's biodiversity. Large portions of this diversity remain undiscovered, moreover the evolutionary processes that bring about this immense diversity wait to be fully understood. In order to study these processes and patterns, the accurate and robust delimitation of evolutionary units is the necessary first step. Species recognition in understudied and unexplored regions containing high levels of diversity is especially challenging due to the difficulty in obtaining ecological and life history data in the field, and in some cases, impossible due to habitat loss. In addition, many of these taxa have undergone recent and rapid speciation. Morphological characters and genetic difference that have commonly been used to circumscribe species often fail to accurately characterize closely related species. Accurate species delimitation may also be obscured by processes such as incomplete lineage sorting and gene flow between sister taxa. Here I analyze what constitutes evolutionary units in a group of diving beetles found in streams throughout the island of New Guinea. I use an integrative set of approaches that are well suited for delineation of recently diverging taxa. I found that results of species validation analyses between sister taxon are highly sensitive to the methods and prior used. I followed an extremely conservative scheme and only accredit hypotheses that are supported across methods.

INTRODUCTION

The robust and accurate delimitation of evolutionary units is the first step that is required for the investigation of most biological questions. Species delimitation is still a developing field that is constantly evolving to accommodate data type and availability. Which species concept is used to delimit species may also largely depend on what data type is available. The general lineage concept (De Queiroz 1999), contends that the only criteria the species category requires is for it to be a separately evolving metapopulation lineage (De Queiroz 2005; Reeves & Richards 2011). Which data types are used to make this distinction are of primary importance. Several authors have recognized that the units diagnosed by multispecies coalescent delimitation methods are simply genetic structure and that other information on the organism's biology is required to meet the criteria of species (Pyron *et al.* 2016a; Sukumaran & Knowles 2017). This necessitates an integrative approach as has been called for by many authors in the pre-genomics era (Will *et al.* 2005). This is especially important for species that have an intermediate level of differentiation i.e. "gray zone" species (De Queiroz 2007) where one or more species delimitation criteria fails e.g. reciprocal monophyly or complete reproductive isolation. Compounding the difficulty of delimiting species in the "gray zone", is the inference of phylogenetic relationship in recently diverged species (Maddison & Knowles 2006). A large proportion of these species may also lack fixed morphological characters which may not have accumulated to a great enough extent for their recognition. In many cases, there is also the lack of variable and informative phylogenetic markers. Additionally, gene trees of closely related species may show conflicting topologies due to incomplete lineage sorting and/or gene flow (Maddison 1997; Wendel & Doyle 1998; Degnan & Rosenberg 2009). Thus, species delimitation should rely upon multiple lines of evidence for accurate and robust delimitation.

For diverse and cryptic tropical insects, where most of the earth's macroscopic species remain to be discovered, obtaining these independent data sources can be a difficult task. Ideally I would want to incorporate morphology, ecology, and genetics to provide robust species delimitation. However, the time investment in the field to study diet choice, behavior, mate choice, phenology and other ecological aspects can be cumbersome when only a few individuals of a species have been found, additionally the taxa under study remain cryptic throughout most of their life history, moreover access to field sites may be very limited. This is particularly true for organisms found in remote tropical regions, as is the focus here. This necessitates the need to sample genomic data and conduct robust analyses to address the questions of gene flow and differentiation, while still integrating information on morphology.

The recent advance in next generation sequencing (NGS) technologies opened up new possibilities for phylogenetic studies of closely related species (Emerson *et al.* 2010; Wagner *et al.* 2013; Eaton & Ree 2013; Hipp *et al.* 2014). In particular Restriction-site-associated DNA sequencing (RAD-seq, Baird *et al.* 2008) allows efficient and cost effective generation of millions of reads throughout the genome of non-model organisms. Consequently, RADseq has been widely used for delimiting species in recently diverged groups (Davey & Blaxter 2010; Jones *et al.* 2013; Nadeau *et al.* 2013; Wagner *et al.* 2013; Eaton & Ree 2013; Leache *et al.* 2014a). Here I aim to integrate genome wide RAD-seq data with morphology to identify species boundaries in the aquatic diving beetle genus *Philaccolilus* from New Guinea.

The development of coalescent-based approaches has also helped to advance molecular species delimitation and are commonly used to validate species hypotheses in a quantitative manner. However, these methods generally assume no gene flow between sister species. This assumption is most likely violated in recently radiating natural systems (Hey &

Pinho 2012). Although these methods claim to be robust, even in systems with low levels of gene flow (Zhang et al. 2011), gene flow may bias species trees estimates (Leache *et al.* 2014b). Therefore, additional approaches based on population genetics theory should be used when studying systems with recently diverging taxa where gene flow may still be present and an estimation of gene flow should be performed to assess the relative importance of gene drift vs gene flow in such systems (Gottscho *et al.* 2017).

The island of New Guinea has one of the most complex geotectonic histories on Earth. The island's geological complexity and rugged topography lead to many isolated environments containing a unique and diverse ensemble of fauna and flora (e.g. Toussaint *et al.* 2014), with some large-scale radiations in these environments (Riedel *et al.* 2013). However, the fauna and flora of the island remains largely understudied. This is especially true for hyperdiverse arthropods, which generally received less attention than charismatic megafauna. Logistical challenges and difficult terrain also contribute to a usually highly patchy sampling regime, which might be a challenge for molecular species delineation approaches due to a lack of intermediate genotypes.

Here, I aim to test the feasibility of multiple species delineation approaches to sustainably and objectively assess hyperdiverse arthropods in this challenging setting. Given that obtaining enough genomic data is less of a challenge I investigate to what extent delimitation methods can accommodate rarity of samples. I use a large Sanger sequencing (SGS) and Next generation sequencing (NGS) dataset. Our test system is a genus of predaceous diving beetles (Dytiscidae, *Philaccolilus*) that is endemic across the entire island of New Guinea. *Philaccolilus* is a small genus, but chosen as it is comparably well studied taxonomically and geographically, and thus manageable for the purpose of this study. Members of this genus inhabit forest streams, some species live along the edges of fast moving rivers in strong current. Species of this genus have distinctive yellow variegated markings, an adaptation for camouflage in sandy/gravelly streams (Balke *et al.* 2000). Present descriptions of species within this genus are based mainly on morphological traits. There are 12 currently described species (Nilsson 2013), and five to six additional species that have been recently discovered (Balke unpublished). Most species seem to have small ranges, e.g. being confined to the Papuan Peninsula, or isolated mountain ranges such as *P. aterrimus*, whereas other morphologically defined species have wider ranges, such as *P. ameliae* across the central highlands and *P. irianensis / P. incognitus* with west-east vicariance along the north coast mountain ranges (Balke *et al.* 2000). *Philaccolilus* species exibit elevational zonation, potentially adapted to different altitudes and stream types (Balke *et al.* 2000). Species seem to be formed in allopatry, as seen by the disjunct distributions between some sister species (Balke *et al.* 2000).

In this study, I utilize an integrative taxonomic approach considering morphological data as well as genetic variation (Sangar sequencing, and Nextera-tagmented, reductively-amplified DNA genotyping). I aim to (1) robustly examine the phylogenetic relationship of *Philaccolilus*; (2) delimit discrete evolutionary units within the phylogeny; (3) estimate demographic parameters of recently diving sister lineages.

MATERIAL AND METHODS

Taxon Sampling

I sampled 89 individuals of *Philaccolilus* from localities across New Guinea. Representing 11 described species of the genus as well as several putative new species from across the range of the genus to provide a preliminary phylogenetic hypothesis (Table S1). One individual

from a closely related genus, *Laccophilus* sp. (Dekai), was included as an outgroup. DNA was extracted from whole beetles with punctured metacoxa, using the DNeasy Blood & Tissue Kit from Qiagen (Hilden, Germany). Vouchers are kept at the Zoological State Collection in Munich as well as at the Museum Zoologicum Bogoriense (Cibinong, West Java, Indonesia).

Sanger Sequencing (SGS)

I sequenced nine gene fragments with an alignment of 5068 base pairs (bps) consisting of fragments from cytochrome oxidase subunit 1 (CO1), Histone 3 (H3), carbamoyl-phosphate synthetase 2 (CAD), Wingless (Wg), RNA polymerase II (RNApol II) (3 non-overlapping fragments), Arginine kinase (AK), Elongation factor 1α (EF1 α). Primers and PCR conditions are listed in Table 1 Sequences were edited using Sequencher 4.10.1 (GeneCodes Corp., Ann Arbor, MI, USA). Sequences were aligned in Mesquite 3.04 (Maddison & Maddison 2008), they were then color-coded by amino acid and checked for stop codons, and finally alignments of all genes were concatenated into a combined matrix. New sequences were deposited in GenBank (accession Nos. LT615409-615637).

nextRAD Sequencing

Representative subsets of samples from each clade are selected for nextRAD (Nextera-tagmented, reductively-amplified DNA) genotyping to collect SNP data DNA samples were sent to SNPsauraus LLC (Eugene, OR) to generate nextRAD libraries and sequencing DNA samples were sent to SNPsaurus (snpsaurus.com) for generation of nextRAD libraries and Illumina sequencing. Genomic DNA was converted into nextRAD libraries as described by Russello *et al.* (2015). Genomic DNA was first fragmented with Nextera reagent (Illumina, San Diego, USA), which also ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled for fragmenting 10 nanograms of genomic DNA. Fragmented DNA was then PCR amplified, with one of the primers matching the adapter and extending nine nucleotides into the genomic DNA with the selective sequence GTGTAGAGC. Therefore, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer were efficiently amplified by PCR. The nextRAD libraries were sequenced on a HiSeq 2500 (University of Oregon, USA).

The genotyping analysis used custom scripts (SNPsaurus) that created a de novo reference from abundant reads, and then mapped all of the reads to the reference with an alignment identity threshold of 93% (BBMap, http://sourceforge.net/projects/bbmap/). Genotype calling was done using Samtools and bcftools (samtools mpileup -gu -Q 10 -t DP,DPR -f ref.fasta -b samples.txt | bcftools call -cv - > genotypes.vcf). The vcf files were converted to PHYLIP format by concatenating the de novo reference and substituting the called genotypes for each sample at the polymorphic positions.

I tested all polymorphic loci for signals of selection using the Bayesian simulation method of Beaumont & Balding (2004) as implemented in BayeScan 2.1. (Foll & Gaggiotti 2008). Analyses were run separately for each cluster studied in detail (described below). I used a prior odds value of 10, with 100,000 iterations and a burn-in of 50,000 iterations. I identified loci that were significant outliers at a *q*-value of 0.20.

Sequencing of the nextRAD library produced a total of 130,005,273 reads from 95 individuals, and the reads collapsed to 40,059 initial loci that were distinct from other loci by an identity threshold of at least 92%. These loci were used as a *de novo* reference for aligning the sequence reads from each sample using bbmap (k=9, slow mode, indel=15, minid=.92). The resulting bam files were converted to a vcf genotype table using samtools mpileup. The

putative variants in the vcf genotype table were then filtered using vcftools to remove variants not present in at least 85%, 95% of the samples and allowing variants with a population frequency of at least 0.05 to reduce artifactual variants. After filtering, the final dataset consists of 5,609 SNPs in 1,726 loci across 90 individuals. The vcf file was converted to phylip format by concatenating the full sequence of each locus and then for each sample substituting the alleles found to create a sample-specific sequence. Thus, it contains the variant and invariant nucleotides for each sample. From the independent SNPs tested for statistical outlier, BayeScan determined that no loci displayed signal of selection in any of the three clusters, therefore all 5,609 SNPs were retained for further analyses.

Phylogenetic analyses

Sanger sequencing (SGS) dataset

I used Bayesian inference (BI) and maximum-likelihood (ML) to reconstruct phylogenetic from the SGS dataset. The partitions and corresponding optimal models of substitution were searched under PartitionFinder 2 (Lanfear *et al.* 2016) using the 'greedy' algorithm, either the 'mrbayes' or 'raxml' set of models and the Akaike Information Criterion corrected (AICc) to compare the fit of the different models.

The BI analyses were performed using MrBayes 3.2.2 (Ronquist *et al.* 2012). I performed two different analyses; one using the substitution models based on the results of PartitionFinder, and one based on the different partitions recovered in PartitionFinder but using reversible-jump MCMC to explore the entire space of substitution models (Huelsenbeck *et al.* 2004). For each analysis, I used two simultaneous and independent runs consisting of eight Metropolis-coupled Markov chain Monte Carlo chains (MCMC, one cold and seven incrementally heated) running 50 million generations, with sampling every 5000 generations to calculate posterior probabilities (PP). In order to check the convergence of the runs, I investigated the split frequencies and Effective SampleSize (ESS) of all the parameters, and plotted the log-likelihood of the samples against the number of generations in Tracer 1.5 (http://BEAST.bio.ed.ac.uk/Tracer). A value of ESS>200 was acknowledged as a good indicator of convergence. All the trees sampled before the log-likelihood plateau were discarded as the burn-in, and the remaining samples were used to generate a 50% majority rule consensus tree.

I also used IQ-TREE (Nguyen *et al.* 2015) as implemented on the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) to look for similarities between the results of the MrBayes and RaxML analysis. The concatenated dataset was left unpartitioned and the best-fit model of substitution was searched using the Auto function on the IQ-TREE web server based on the AICc. I performed 1,000 ultrafast bootstrap replicates (Minh *et al.* 2013) to investigate nodal support across the topology.

The ML analyses were conducted with the best partitioning scheme selected in PartitionFinder 2 (Lanfear *et al.* 2012) using RAxML AQ11(Stamatakis 2006). I used the AutoFC option in RAxML to calculate the optimal number of Bootstrap replicates (BS) indicating the level of support at each node. A calculated PP≥0.95 or a BS≥70 was considered to indicate strong support for a given clade (Hillis & Bull 1993; Erixon *et al.* 2003).

NextRAD dataset

Phylogenetic under the multispecies coalescent (MSC). For each of the 1,726 loci, I conducted a ML analyses under the GTR+G model in RaxML. I then constructed a gene tree for each one of the loci, using RAxML 8.0.19 (Stamatakis 2014) by conducting 20 searches of

the data to select the best ML tree. The 'best' tree for each locus were then used to construct the multispecies coalescent history using multiple individuals per-species when available in ASTRAL-II (Mirarab *et al.* 2014). I assessed support by conducting 200 bootstrap replicates per gene.

I used a second method consistent with the MSC to assess topological consistency between methods, SVDquartets (Chifman & Kubatko 2014, 2015). I used the evalQuartets=all option to evaluate all quartets. I assessed support by conducting 200 bootstrap replicates.

Concatenated phylogenetic analyses. I conducted phylogenetic analyses by generating consensus sequences of all individuals of each species in Sequencher 5.0.1 (GeneCodes Corp., Ann Arbor, MI, USA). The best partitioning scheme was selected in PartitionFinder 2 (Lanfear *et al.* 2016) using the 'greedy' algorithm. In the gene tree analyses, character sets were defined according to the number of phylogenetically informative sites, giving a total of 10 character sets. The ML analyses were performed using RAxML 8.0.19 (Stamatakis 2014) with a GTR+G model for each partition, conducting 20 independent tree searches for the best ML tree. I assessed support for the best ML topology by performing nonparametric bootstrapping using the autoMRE option in RAxML.

I also performed partitioned concatenated Bayesian inference analyses with MrBayes 3.2 (Ronquist *et al.* 2012). The best partitioning scheme and corresponding models of substitution were searched in PartitionFinder 2. The analysis consisted of two independent runs of four Markov chains Monte Carlo (MCMC) running 20 million generations with tree and parameter sampling every 1000 generations. After removing 30% of the posterior samples as burn-in, I generated a 50% majority-rule consensus topology. The posterior values were examined in Tracer v1.0.6 (Rambaut *et al.* 2014) to ensure that the Effective Sample Size (ESS) of each parameter was sufficiently sampled.

Genetic species delimitation

To identify robust and geographically defined genetic clusters and to estimate their phylogenetic relationship, I employed two categories of methods for species discovery and species validation. **1**, in order to identify major genetic clusters and evaluated genetic exchange between sister taxa, I used the Bayesian clustering approach implemented in the program STRUCTURE 2.3.4 (Pritchard *et al.* 2000), I use a hierarchical approach to identify various level of genetic clustering and visualize the degree of admixture. **2**, I evaluated the distinctiveness of putative species clusters using two coalescent based species validation methods. I used the program BPP (Bayesian Phylogenetics and Phylogeography, Rannala & Yang 2003; Yang & Rannala 2010, 2014) which simultaneously estimates species trees and species delimitation models. In addition, I used the program bPTP which distinguish speciation and coalescent processes based on branch-length and topology of a phylogeny. **3,** I estimated demographic parameters using the isolation-with-migration model implemented in the program GPhoCS (Gronau *et al.* 2011) to infer how demographic histories affect the distinctiveness of each clade.

1) Identification of genetic cluster

I used the Bayesian clustering approach implemented in the program STRUCTURE 2.3.4 (Pritchard *et al.* 2000). I ran 10 replicates, each using a burn-in length of 100,000 and a run length of 1,000,000 steps, with the admixture and the correlated allele frequencies models without using prior population information (geographic sampling location). I varied the number of clusters (*K*) from 1 to 18. The broad scale number of clusters (distinct clades) was

initially determined by examining both the posterior probabilities of the data for each K and the ΔK estimator described by Evanno *et al.* (2005) as calculated in Structure Harvester (Earl & vonHoldt 2012). Results for the identified optimal values of *K* were summarized using CLUMPP ver. 1.1 (Jakobsson & Rosenberg 2007) using 1000 permutations and the LargeKGreedy algorithm; the result is then plotted using DISTRUCT ver. 1.1 (Rosenberg 2004). To detect subdivision that may be overlooked within each cluster (for example, presence of closely related species within clades) and/or to assess the degrees of genetic exchange between geographic populations, I subsequently conducted STRUCTURE analyses for each subclade, as in Gowen *et al.* (2014).

2) Coalescent species validation method

In light of our STRUCTURE results I first defined potential genetic units and assign individuals to these units. I then used BP&P and bPTP to validate these potential species.

I used the program BPP v. 3.3 (Rannala & Yang 2003; Yang & Rannala 2010, 2014) to conduct a coalescent-based species-validation methods in order to evaluate the distinctiveness of the putative species clusters. Initially, I used algorithm (A11), which has the ability to simultaneously estimate a species tree and species delimitation models. However, I recovered implausible results from the A11 runs based on our knowledge from our phylogenetic and STRUCTURE analyses due to issues of mixing. Therefore, I used algorithm A10 which only estimates delimitation models given a provided guide trees based on our species trees obtained from ASTRAL. I used the Leaché & Fujita (2010), approach where I used three different combination of priors for the root age and ancestral population size to examine if our delimitations were sensitive to different priors. The root age and ancestral population size are both assigned a gamma prior distribution Γ (α, β). I used three different combinations of these: large ancestral populations and old divergences, small ancestral population size and recent divergences, large ancestral population sizes and recent divergences, see Table 2 for the specific settings. Within each of these subsets I pruned the data to include more than two individuals and at least one phylogenetically informative site (segregating site) to eliminate uninformative loci.

I used a second validation approach, the Bayesian implementation of the bPTP (Zhang *et al.* 2013) to infer genetic clades based on our molecular phylogenies. The PTP model allows us to distinguish speciation and coalescent processes based on the branch-lengths and topologies of our input phylogenies. I compared results based on four different phylogenies (i) MrBayes phylogenetic reconstruction using our nextRAD dataset, (ii) RAxML phylogenetic reconstruction using our nextRAD dataset, (iii) MrBayes phylogenetic reconstruction using our SGS dataset, and (iv) RAxML phylogenetic reconstruction using our SGS dataset. The analyses were conducted on the web server for bPTP (http://species.h-its.org/ptp/) using each of the four topologies. Each analysis consisted of 500 000 generations, with a thinning every 100 generations and a burn–in of 25%.

3) Isolation-with-migration model.

To infer how the demographic history of each clade affects the distinctiveness of taxa within these groups I estimated migration rates (m), divergence times (τ), and effective population sizes (θ) using the program G-PhoCS (Generalized Phylogenetic Coalescent Sampler) v1.2.3 (Gronau *et al.* 2011). The prior probability distributions for all models assumed a gamma distribution. For τ and θ I used α = 1, β = 10,000, For *m* I used α = 0.002, β = 0.00001. All finetune parameters were set automatically. I ran 1 million MCMC steps, discarding the first

10% as burnin. Tracer v.1.6.0 was used to ensure that all ESS values were >200 and that the combined runs converged on the same posterior distributions for all parameters.

In order to estimated μ (substitutions/site/year) using the equation $\mu = T/\tau$ (where T is the divergence time for each of the clades), I needed to estimate T for each clade. I employed molecular dating analyses to infer the time of population divergence based on CO1 data. I used a relaxed molecular clock in BEAST 2.4.4 (Drummond *et al.* 2012). I used COI codon positions as our character sets and used PartitionFinder version 2.1.1 (Lanfear *et al.* 2012) to find the 'best' partitioning strategy. I used the YULE tree model and pruned each population to the individual with the most complete data. This was done because the YULE model assumes that each terminal is a species/monophyletic coalescing unit. I used the uncorrelated relaxed clock model that has a lognormal distribution of rates. I set a uniform prior for the mitochondrial clock rate between 0.0145–0.0195, as these rates have been reported in dytiscid beetles (Toussaint *et al.* 2014). I ran four different independent runs of BEAST for 50 million generations sampling every 5000 generations. I then examined each run in Tracer and removed the burn-in (10%) and combined each run in Logcombiner 1.8.0 (Drummond *et al.* 2012). The maximum credibility tree was generated in TreeAnnotator 1.8.0 (Drummond *et al.* 2012).

I used $\theta = 4N\mu g$ (N= number of diploid individuals; g= average generation time, in our case 0.5 year) to estimate effective population sizes; I used $M = m\mu g$ to convert migration rates (*m*) into per-generation migration rate.

According to Wright's island model (Wright 1931), when $2NM > 1$, the rate of gene flow is significant relative to the rate of gene drift, the genetic makeup of the sink population mimics that of the source, and therefore, divergence is not likely to take place (Wright 1931). Although Wright's model is problematic, for example it assumes no natural selection (Wright 1931), I used $2NM = 1$ as a reasonable as a cutoff for delimiting species as it reflects a sensible, approximate indicator of reproductive isolation (Gottscho *et al.* 2017). Following Gottscho *et al.* 2017, if mean $2NM > 1$, I rejected the hypothesis that the taxa is a distinct species; conversely, if mean $2NM < 1$, I confirmed the hypothesis.

RESULTS

Phylogenetic analyses

nextRAD dataset: I resolve a robust phylogeny using the NGS data. Our BI and ML analyses yielded identical phylogenetic trees (Fig.1). The ingroup relationships among species are well supported. The morphospecies *P. bellissimus*, *P. bicinctus*, *P. mekus*. *P. bacchusi*, and *P. mas* remain distinct lineages. *Philaccolilus. mas* is nested within the *P. ameliae* complex, consisting of morphospecies collected from different localities. Five closely related clades are recovered: (A) consisting of *P. ameliae ameliae* population from Ok Sibil as well as a distinct subspecies (*P. ameliae weylandensis*) collected from the Weyland mountains. (B) *P. ameliae ameliae* populations from the Foja mountains and the Sandaun province. (C) *P. ameliae ameliae* populations samples from three localities in the Bird's Head peninsular. (E) *P. black* and *P. band*, two morphologically distinct putative species.

MSC analyses recovers, with strong support (Figure 2 & 3) a similar topology to the concatenated analyses, recovering monophyly for all five monophyletic clades mentioned above. However, there is substantial difference along the backbone of the tree. The analyses from *SVDquartets* showed a similar topology to of ASTRAL with two main differences. One,

that *P. bacchusi* is sister to the *P. ameliae* species group. Two, that *P. bicinctus* is sister to *P. mekus* and remainig *Philaccolilus*. Bootstrap support was high except for the *P. bicinctus*, *P. mas*, and *P. bellissimus* nodes.

Sanger sequencing dataset: Phylogenetic analyses based on the Sanger sequencing dataset of 5068 bps implementing different approaches and parameter choices (IQtree, RAxML, MrBayes, MrBayes reversible -jump) produced conflicting topologies (Fig S1). The backbone of the phylogeny has generally weak nodal support. Five monophyletic clades of closely related taxa are recovered (described in detail below).

Genetic species delimitation

Genetic clustering

The population structure analysis provides a powerful approach for exploring the relationships among multiple individuals. STRUCTURE analyses including all individuals is congruent with our phylogenetic results. It reveals the same distinct units described above, represented by ten color blocks in figure 4. Subsequent analyses reveal high levels of admixture between populations within clusters B, C, and E. In cluster A, *P. ameliea* from Ok Sibil and *P. ameliae weylandensis* form two distinct genetic populations (or putative species). In cluster D, there is a moderate level of genetic exchange between the two morphospecies *P. iratoi* and *P. ramuensis.* The single individual of *P. alderbert* and *P. irianensis* share the same genetic grouping, this is confirmed by further structure analysis including only the five individuals (not shown). Further analyses will focus on the five clusters A through E.

Coalescent species validation method

The results of the different species delimitation analyses are summarized in (Figure 5).
I found that the results from the A11 algorithms in BPP are extremely sensitive to models and priors used. BPP has several different algorithms for delimiting species, one can use a combination of fixed species groupings or fixed trees or estimate both simultaneously. From preliminary analyses, I found that BPP gave conflicting results when the species tree was not fixed as well as what phylogenetic scale the analyses were conducted on (all 18 putative species versus small clades). I therefore used a fixed species tree topology to eliminate this variable as I already had a species tree topology from ASTRAL-II. I selected terminals based on a maximally split species concepts from the STRUCTURE results plus morphology for those not detected as separate species by STRUCTURE. From this final analysis, the three sets of priors tested in BPP still resulted in discrepancies between priors within certain clades. For example, *P. black* and *P. band* (clade E) are considered two different units under one model (big ancestral population size, deep divergence in this case) but are clumped under the other two models; the same is true for the sister taxon *P. adelbert* & *P. irianensis* (clade D) and *P. ameliae* from the Weyland mountains & Oksibil (clade A).

The bPTP analyses from both NGS and SGS data using RAxML and MrBayes topologies resulted in the same delimitation with varying degrees of support. bPTP results based on SGS data, in general, resulted in weaker support when compared to nextRAD dataset. There are discrepancies between results generated from the two datasets. SGS results clumps *P. black* and *P. band* into one unit while NGS results keep the species separate. NGS results separate clade D into two species (P. *ramuensis* + *P. iratoi and P. adelbert* + *P. irianensis*) while SGS results clump all four morphospecies into one taxa.

Isolation-with-migration
Effective population size (N) and, migration rate (2NM) based on *m* and θ values obtained from GPhoCS are presented in Table 3. Analyses for clade C, consisting of three *P. ameliae* populations on the Bird's Head peninsular, was never able to converge unless migration bands are disregarded. Therefore, no results from this clade are reported.
High gene flow (2Nm>1) is observed between; the two *P. ameliae* populations found in Foja and Sandaun Mtns., between *P. irianensis* and *P. aldelbert*, and between *P. black* and *P. band*.

Discussion

The main objective of this study was to delineate independent evolutionary units within *Philaccolilus*. I first utilized a hierarchical Bayesian clustering approach to illuminate the major genetic groupings and visualize the amount of admixture within each lineage; I treated each population as a potential species in the MCS delimitation analyses. Not surprisingly the sister pairs with the most admixture where also the most inconsistently delimited by the MSC. I compared two species validation methods (BPP and bPTP) and different priors within each, I find high levels of discrepancy in the results, indicating that the dataset is highly sensitive to priors and methods used. In our study, all five putative species clades A-E are supported by all three sets of priors tested. However, there are discrepancies in species delimitation results within clades based on different priors. This discrepancy is also observed when I compare results from BBP to that of bPTP. Within the bPTP analyses, differences between datasets (SGD vs NGS) also resulted in different delimitation schemes within clades.

Knowles & Carstens 2007 suggested that coalescent-based species delimitation methods (such as BBP) may over-lump species when divergence is recent. It is also possible that the RAD loci used in our study have not accumulated enough informative differences in recently diverged species, such as would be the result from rapid morphological change resulting in speciation. Moreover, BBP also tends to over-lump taxa if there is gene flow between sister species (Zhang *et al.* 2011). On the other hand, O'Meara 2010 suggested that coalescent methods may over-split taxa if there is population subdivision within a lineage.

Carstens *et al.* 2013 argued that delimitation of species should incorporate an integrative range of species delimitation analyses and only accredit hypotheses that are supported across methods. Following this extremely conservative scheme, Table 4 summarizes the support for each taxon based on the different analysis. I take a conservative stance, such that unless all species discovery and validation methods support the recognition of a taxa as a distinct lineage, I retain these metapopulations as one entity.

Philaccolilus. ameliae as currently described is paraphyletic in respect to *P. mas*. Table 4 makes it clear that the currently described morphospecies should be divided into at least three separate species: One consisting of the *P. ameliae* population from OkSibil and the *P. a weylandensis* subspecies found in Nabire (clade A), although there is limited gene exchange between these two taxa as shown by STRUCTURE as well as low migration rate estimated in GPhoCS, the split is generally not supported by BBP and bPTP validation methods (except for BBP M1, however posterior probability is less than 0.50). The second putative species includes *P. ameliae* found in Foja as well as those individuals found in Sandaun (i.e. clade B), this is supported by all analyses performed. The third putative species includes the three populations found on the Bird's Head peninsula (i.e. clade C). Although I was unable to directly estimate gene flow using GPhoCS due to computational limitations, all other analyses confirmed that the three populations are genetically extremely similar.

Within clade B, all analyses indicate that *P. ramuensis* and *P. "iratoi"* represent distinct species. *P. adlebert* and *P. irianensis* appears to be genetically similar and may be grouped into one evolutionary unit.

Philaccolilus. black and *P. band* are both found in the Bird's Head peninsular. All of our *P. black* samples were collected in Testega whereas *P. band* are from Tamrau. Although the two locations are only about 50 km apart, *P. black* and *P. band* individuals have consistently different color patterns and body shape (Balke unpublished). However, their genital morphology is indistinguishable. Despite this morphological difference, STRUCTURE and GPhoCS analyses indicates that there is a high level of gene flow between the two taxa. However, this separation is not supported by four out of seven of the species validation methods used.

Given the difficulties in sampling in remote areas, there are likely many new species and/or populations to be sampled. Moreover, considering that gene flow is such an important factor when determining a "gray zone" species' status I question if the MSC methods are the best way to delimit species when applied to patchy sampling. Certainly, they are an objective assessment which definitely improves the delimitation process; but without morphological or ecological differences, one should be cautious when designating new species based solely on genetic criteria as others have suggested (e.g. Carstens *et al.* 2013). However, obtaining consistency between methods where there are few samples with some gene flow is problematic as some methods give conflicting results as in our study with BPP and bPTP, others still are too computationally expensive given large genomic datasets; but at least one can look for consistency and argue for the need to study/sample more thoroughly for the taxa in question. While it is challenging to obtain samples, gathering enough genetic information from each is less of a problem, however the issue of exploration and natural history is not going away (Lim *et al.* 2012), and thus one should not be dissatisfied when reaching the conclusion that we still can't decide with certainty if these entities should be lumped or split. I anticipate that this conclusion should be reached more often especially in diverse arthropod groups. While the need for continued theoretical and methodological exploration is greater than ever the same cavear is true for exploration and natural history studies in remote regions of the world.

REFERENCES

Arribas P, Velasco J, Abellán P *et al.* (2012) Dispersal ability rather than ecological tolerance drives differences in range size between lentic and lotic water beetles (Coleoptera: Hydrophilidae). *Journal of Biogeography*, **39**, 984–994.

Baird N, Etter P, Atwood T, Currey M (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, **3**, e3376.

Balke M, Jäch MA, Hendrich L (2004) Coleoptera. In: *Freshwater invertebrates of the Malaysian region* (eds Yule CM, Yong HS), pp. 555–609. Akademi Sains Malaysia, Kuala Lumpur, Malaysia.

Balke M, Larson D, Hendrich L, Konyorah E (2000) A revision of the New Guinea water beetle genus Philaccolilus Guignot, stat.n. (Coleoptera, Dytiscidae). *Deutsche Entomologische Zeitschrift*, **47**, 29–50.

Balke M, Ribera I, Hendrich L *et al.* (2009) New Guinea highland origin of a widespread arthropod supertramp. *Proceedings of the Royal Society of London B: Biological Sciences*.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, Population genetics software for Windows TM.

Bilton DT, Freeland JR, Okamura B (2001) Dispersal in Freshwater Invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.

Bohonak A (1999) Dispersal, gene flow, and population structure. *Quarterly review of biology*, 21–45.

Bowler DE, Benton TG (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological reviews of the Cambridge Philosophical Society*, **80**, 205–25.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.

Chifman J, Kubatko L (2015) Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, **374**, 35–47.

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–340.

Deiner K, Lemmon AR, Mack AL, Fleischer RC, Dumbacher JP (2011) A Passerine Bird's Evolution Corroborates the Geologic History of the Island of New Guinea (B Schierwater, Ed,). *PLoS ONE*, **6**, e19479.

Denno R, Olmstead K, McCLoud E (1989) Reproductive cost of flight capability: a comparison of life history traits in wing dimorphic planthoppers. *Ecological Entomology*, **14**, 31–44.

Dijkstra K-DB, Monaghan MT, Pauls SU (2014) Freshwater Biodiversity and Aquatic Insect Diversification. *Annual Review of Entomology*, **59**, 143–163.

Dobson M, Frid C (1998) Ecology of aquatic ecosystems. *Essex: Addison Wesley Longman Limited*.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.
Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.
Edwards JS, Sugg P (1993) Arthropod fallout as a resource in the recolonization of Mount St. Helens. *Ecology*, **74**, 954–958.
Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, **107**, 16196–16200.
Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic biology*, **52**, 665–673.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplo- types: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.
Gaston KJ (2000) Global patterns in biodiversity. *Nature*, **405**, 220–227.
Georges A, Spencer R, Welsh M *et al.* (2011) Application of the precautionary principle to taxa of uncertain status: the case of the Bellinger River turtle. *Endangered Species Research*, **14**, 127–134.
Gottscho AD, Wood DA, Vandergast AG *et al.* (2017) Lineage diversification of fringe-toed lizards (Phrynosomatidae: Uma notata complex) in the Colorado Desert: Delimiting species in the presence of gene flow. *Molecular Phylogenetics and Evolution*, **106**, 103–117.
Gowen FC, Maley JM, Cicero C *et al.* (2014) Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC evolutionary biology*, **14**, 135.
Grewe Y, Hof C, Dehling DM, Brandl R, Brändle M (2013) Recent range shifts of European dragonflies provide support for an inverse relationship between habitat predictability and dispersal. *Global Ecology and Biogeography*, **22**, 403–409.
Gronau I, Hubisz M, Gulko B, Danko C, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics*.
Hey J, Pinho C (2012) Population genetics and objectivity in species diagnosis. *Evolution*, **66**, 1413–1429.
Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**, 182–192.
Hipp AL, Eaton DAR, Cavender-Bares J *et al.* (2014) A framework phylogeny of the American oak clade based on sequenced RAD data (S Joly, Ed,). *PLoS ONE*, **9**, e93975.

Hjalmarsson A, Bergsten J, Monaghan M (2014) Dispersal is linked to habitat use in 59 species of water beetles (Coleoptera: Adephaga) on Madagascar. *Ecography*, **38.7**, 732–739.

Hof C, Brändle M, Dehling DM *et al.* (2012) Habitat stability affects dispersal and the ability to track climate change. *Biology Letters*, **8**.

Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian Phylogenetic Model Selection Using Reversible Jump Markov Chain Monte Carlo. *Molecular Biology and Evolution*, **21**, 1123–1133.

Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Janda M, Matos-Marav, P, Borovanska M *et al.* (2016) Phylogeny and population genetic structure of the ant genus Acropyga (Hymenoptera : Formicidae) in Papua New Guinea. *Invertebrate Systematics*, **30**, 28.

Jones J, Fan S, Franchini P (2013) The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular ...*, **22**, 2986–3001.

Knowles LL, Carstens B (2007) Delimiting Species without Monophyletic Gene Trees. *Systematic Biology*, **56**, 887–895.

Korfiatis KJ, Stamou GP (1999) Habitat Templets and the Changing Worldview of Ecology. *Biology and Philosophy*, **14**, 375–393.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution*, msw260.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295–302.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society of London B: Biological Sciences*.

Leache AD, Fujita MK, Minin VN, Bouckaert RR (2014a) Species delimitation using genome-wide SNP Data. *Systematic Biology*, **63**, 534–542.

Leache AD, Harris RB, Rannala B, Yang Z (2014b) The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Systematic Biology*, **63**, 17–30.

Letsch H, Gottsberger B, Ware JL (2016) Not going with the flow: a comprehensive time-calibrated phylogeny of dragonflies (Anisoptera: Odonata: Insecta) provides evidence for the role of lentic habitats on diversification. *Molecular Ecology*, **25**, 1340–1353.

Ligges U, Martin M (2003) Scatterplot3d – an R package for Visualizing Multivariate Data. *Cran*, **8**, 1–36.

Lim GS, Balke M, Meier R (2012) Determining Species Boundaries in a World Full of Rarity: Singletons, Species Delimitation Methods. *Systematic Biology*, **61**, 165–169.

Maddison WP (1997) Gene trees in species trees. *Systematic biology*, **46**, 523–536.

Maddison WP, Knowles LL (2006) Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology*, **55**, 21–30.

Maddison WP, Maddison DR (2008) Mesquite: A modular system for evolutionary analysis.

Evolution, **62**, 1103–1118.
Marten A, Brandle M, Brandle R (2006) Habitat type predicts genetic population differentiation in freshwater invertebrates. *Molecular Ecology*, **15**, 2643–2651.
Minh BQ, Nguyen MAT, Von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, **30**, 1188–1195.
Mirarab S, Reaz R, Bayzid MS *et al.* (2014) ASTRAL: Genome-scale coalescent-based species tree estimation. In: *Bioinformatics*, pp. i541–i548. Oxford University Press.
Monaghan MT, Gattolliat J-L, Sartori M *et al.* (2005) Trans-oceanic and endemic origins of the small minnow mayflies (Ephemeroptera, Baetidae) of Madagascar. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**.
Myers EA, Rodríguez-Robles JA, Denardo DF *et al.* (2013) Multilocus phylogeographic assessment of the California Mountain Kingsnake (Lampropeltis zonata) suggests alternative patterns of diversification for the California Floristic Province. *Molecular Ecology*, **22**, 5418–5429.
Nadeau NJ, Martin SH, Kozak KM *et al.* (2013) Genome-wide patterns of divergence and gene flow across a butterfly radiation. In: *Molecular Ecology*, pp. 814–826.
Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.
Nilsson AN (2013) *A World Catalogue of the Family Dytiscidae , or the Diving Beetles (Coleoptera , Adephaga)*.
Papadopoulou A, Anastasiou I (2009) Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: the effect of dispersal ability and habitat preference. *Molecular Ecology*, **18**, 2503–2517.
Papadopoulou A, Bergsten J, Fujisawa T *et al.* (2008) Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 2987–2996.
Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Pyron RA, Hsieh FW, Lemmon AR, Lemmon EM, Hendry CR (2016a) Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (Storeria). *Zoological Journal of the Linnean Society*, **177**, 937–949.
Pyron RA, Hsieh FW, Lemmon AR, Lemmon EM, Hendry CR (2016b) Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). *Zoological Journal of the Linnean Society*, **177**, 937–949.
De Queiroz K (1999) *The general lineage concept of species and the defining properties of the species*. MIT Press.
De Queiroz K (2005) A Unified Concept of Species and Its Consequences for the Future of Taxonomy. *Proceedings of the California Academy of Sciences. I)*, **4**.
De Queiroz K (2007) Species Concepts and Species Delimitation. *Systematic Biology*, **56**, 879–886.
Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer | BEAST.
Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral

population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–56.
Reeves PA, Richards CM (2011) Species Delimitation under the General Lineage Concept: An Empirical Example Using Wild North American Hops (Cannabaceae: Humulus lupulus). *Systematic Biology*, **60**, 45–59.
Ribera I (2008) Habitat constraints and the generation of diversity in freshwater macroinvertebrates. In: *Aquatic insects: challenges to populations* (eds Lancaster J, Briers RA), pp. 289–311. CABI, Wallingford.
Ribera I, Dolédec S, Downie I, Foster G (2001) Effect of land disturbance and stress on species traits of ground beetle assemblages. *Ecology*, **82**, 1112–1129.
Ribera I, Foster G, Vogler A (2003) Does habitat use explain large scale species richness patterns of aquatic beetles in Europe? *Ecography*, **26**, 145–152.
Ribera I, Vogler P (2000) Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera. *Biological Journal of the Linnean Society*, **71**, 33–52.
Ribera I, Vogler A (2004) Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, **13**, 179–193.
Riedel A, Sagata K, Surbakti S, Rene Tänzler, Michael Balke (2013) One hundred and one new species of Trigonopterus weevils from New Guinea. *ZooKeys*, 1–150.
Roff D (1986) The evolution of wing dimorphism in insects. *Evolution*, **40**, 1009–1020.
Roff D (1994) Habitat persistence and the evolution of wing dimorphism in insects. *American Naturalist*, **144**, 772–798.
Ronquist F, Teslenko M, Van Der Mark P *et al.* (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ*, **3**, e1106.
Short AEZ, Caterino MS (2009) On the validity of habitat as a predictor of genetic structure in aquatic systems: A comparative study using California water beetles. *Molecular Ecology*, **18**, 403–414.
Slatkin M (1985) Gene flow in natural populations. *Annual review of ecology and systematics*, 393–430.
Southwood T (1977) Habitat, the templet for ecological strategies? *The Journal of Animal Ecology*, **64**, 337–365.
Southwood TRE (1988) Tactics, strategies and templets. *Oikos*, **52**, 3–18.
Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 1607–1612.
Toussaint EFA, Hall R, Monaghan MT *et al.* (2014) The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature communications*, **5**, 4001.
Toussaint EFA, Hendrich L, Shaverdo H, Balke M (2015) Mosaic patterns of diversification dynamics following the colonization of Melanesian islands. *Scientific Reports*, **5**, 16016.
Toussaint EFA, Sagata K, Surbakti S, Hendrich L, Balke M (2013) Australasian sky islands

act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. *Ecology and Evolution*, **3**, 1031–1049.
Unmack PJ, Allen GR, Johnson JB (2013) Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. *Molecular Phylogenetics and Evolution*, **67**, 15–27.
Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. In: *Molecular Ecology*, pp. 787–798.
Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.
Wendel J, Doyle J (1998) Phylogenetic incongruence: window into genome history and molecular evolution. *Molecular systematics of plants II*, 265–296.
Will K, Mishler B, Wheeler Q (2005) The Perils of DNA Barcoding and the Need for Integrative Taxonomy. *Systematic Biology*, **54**, 844–851.
Wright S (1931) Evolution in Mendelian populations. *Genetics*.
Yang Z (2015) The BPP program for species tree estimation and species delimitation. *Current Zoology*, **61**, 854–865.
Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264–9269.
Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple Loci. *Molecular biology and evolution*, **31**, 3125–35.
Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, **29**, 2869–2876.
Zhang C, Zhang D-X, Zhu T, Yang Z (2011) Evaluation of a Bayesian Coalescent Method of Species Delimitation. *Systematic Biology*, **60**, 747–761.

TABLES

Table 1: Primers and PCR conditions.

Gene	Location	Primer	Direction	Sequence(5'-->3')	PCR protocol	References
Cytochromecoxidase1	Mitochondrial	Jerry	Forward	CAACATTTATTTTGATTTTTTGG	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Simonetal.,1994
		Pat	Reverse	TCCAATGCACTAATCTGCCATATTA		Simonetal.,1994
Hisone3	Nuclear	H3aF	Forward	ATGGCTCGTACCAAGCAGACGGC	3'96°C–35x(30s94°C–1'52°C–1'72°C)–10'72°C	Colganetal.,1998
		H3aR	Reverse	ATATCCTTGGGCATGATGGTGAC		Colganetal.,1998
carbamoyl-phosphatesynthetaseII_Fragment3	Nuclear	CD821F	Forward	AGCACGAAAATHGGNAGYTCNATGAARAG	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Wild&Maddison,2008
(CAD3)		CD1098R2	Reverse	GCTATGTTGTTNGGNAGYTGDCCNCCCAT		Wild&Maddison,2008
Wingless(Wg)	Nuclear	Wg550F	Forward	ATGCGTCAGGARTGYAARTGYCAYGGYATGTC	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Wild&Maddison,2008
		WgAbRZ	Reverse	CACTTNACYTCRCARCACCARTG		Wild&Maddison,2008
RNAPolymeraseII_Fragment1	Nuclear	PL527F	Forward	AAYAAACCVGTYATGGGTATTGTRCA	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Wild&Maddison,2008
		PL758R	Reverse	ACGACCATAGCCTTBAGRTTRTTRTAYTC		Wild&Maddison,2008
RNAPolymeraseII_Fragment2	Nuclear	PL709F	Forward	GTCATAGAGGTAATCCARAARGCNCAYAAYATGGA	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Wild&Maddison,2008
		PL982R	Reverse	AARATYTTYTGYACRTTCCARATCAT		Wild&Maddison,2008
RNAPolymeraseII_Fragment3	Nuclear	PL859F	Forward	CGTCTGATCAAGGCTATGGARTCNGTNATGGT	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Wild&Maddison,2008
		PL1097R	Reverse	CCAGCGAAGTGGAAVGTRTTNAGBGTCATYTG		Wild&Maddison,2008
ArginineKinase(AK)	Nuclear	AK183F	Forward	GATTCTGGAGTCGGNATYTAYGCNCCYGAYGC	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Wild&Maddison,2008
		AK939R	Reverse	GCCNCCYTCRGCYTCRGTGTGYTC		Wild&Maddison,2008
Elongation Factor 1α (EF1α)	Nuclear	For3	Forward	GGYGACAAYGTTGGTTTYAAY	3'96°C–35x(30s94°C–1'48°C–1'72°C)–10'72°C	Danforth et al. (1999)
		Cho10	Reverse	ACRGCVACKGTYTGHCKCATGTC		Danforth et al. (1999)

Table 2: Species validation methods used.

	Program	**Dataset used**	**BPP Prior / bPTP input topology**
Model1 (M1)	BBP (A10)	nextRAD	Big ancestral population size, deep divergence (BD): Gθ(1,10) ; Gτ(1,10)
Model 2 (M2)	BPP (A10)	nextRAD	Small ancestral population size, shallow divergent (SS): Gθ(2, 2000) ; Gτ(2, 2000)
Model 3 (M3)	BPP (A10)	nextRAD	Big ancestral population size, Shallow divergent (BS): Gθ(1,10) ; Gτ(2,2000)
Model 4 (M4)	bPTP	nextRAD	RAxML topology
Model 5 (M5)	bPTP	nextRAD	MrBayes topology
Model 6 (M6)	bPTP	Sanger	RAxML topology
Model 7 (M7)	bPTP	Sanger	MrBayes topology

Table 3: Results of G-PhoCS models. N = effective population size , T = divergence time (years), 2NM = effective population migration rates.

Clade A

Parameter	Mean
N_*P.ameliae weylandensis*(OkSibil)	6343014
N_*P.ameliae weylandensis* (Nabire)	2273453
N_*P.a.weylandenesis* (OkSibi1 + Nabire)	23365594
T_*P.a.weylandenesis* (OkSibi1 + Nabire)*	2009700
2Nm_*P.a.weylandensis*(OkSibil)->(Nabire)	0
2Nm_*P.a.weylandensis*(Nabire)->(OkSibil)	0.003

Clade B

Parameter	Mean
N_*P. ameliae* (Foja)	4288190
N_*P. ameliae* (Sandaun)	969732
N_*P. ameliae* (Foja+Sandaun)	67498
T_*P. ameliae* (Foja+Sandaun)*	961700
2Nm_*P. ameliae(*Foja)->*P. ameliae* (Sandaun)	1
2Nm_*P. ameliae(*Sandaun)->*P. ameliae* (Foja)	1

Clade D

Parameter	Mean
N_*P. iratoi*	256997
N_*P. ramuensis*	39146
N_*P. aldelbert*	595020
N_*P. irianensis*	1649850
N_*P. irianensis* + *P. aldelbert*	125267
N_*P.ramuensis* + *P. iratoi*	1661780
N_*P. irianensis* + *P. aldelbert* + *P.ramuensis* + *P. iratoi*	32303059
T_*P. irianensis* + *P. aldelbert* + *P.ramuensis* + *P. iratoi**	3299700
2Nm_*P.irianensis->P.aldalbert*	1
2Nm_*P.aldelbert*-> *P.irianensis*	2
2Nm_*P. ramuensis*-> *P.iratoi*	0
2Nm_iratoi->ramuensis	0

Clade E

Parameter	Mean
N_P.band	8237527
N_P.black	4855401
N_*P.band* + *P.black*	155433683
T_*P.band* + *P.black**	5316600
2Nm_*P.band->P.black*	1
2Nm_*P.black->P.band*	2

*T (time of divergence) estimates obtained from BEAST analyses

Table 4: Summary of results of species validation approaches. 'Yes' indicates that the taxon (columns) was supported by the analysis (rows) by more than 0.70 posterior probability, while 'no' indicates that it was not.

	Population Clustering	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**2NM<1**	**Evolutionary Units**
P. band + *P. black*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	***P. "BLACK&BAND"***
P. black	No	No	Yes	Yes	Yes	Yes	No	NA	No	
P. band	No	No	Yes	Yes	Yes	Yes	No	NA	No	
P.ramuensis + *P. iratoi*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	**x**
P.ramuensis	Yes	Yes	Yes	NA	NA	NA	NA	NA	Yes	***P.ramuensis***
P. iratoi	Yes	Yes	Yes	NA	NA	NA	NA	NA	Yes	***P. iratoi***
P. adelbert + *P. irianensis*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	***P. "ADELBERT& IRIANENSIS"***
P. adelbert	No	No	NA	No	NA	No	NA	NA	No	
P. irianensis	No	No	NA	No	NA	No	NA	NA	No	
P. ameliea ameliae (Foja + Sandaun)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	***P. "AMELIAEAMELIAE"***
P. ameliea ameliae (Foja)	No	No	No	No	NA	NA	NA	NA	No	
P. ameliea ameliae (Sandaun)	No	No	No	No	NA	NA	NA	NA	No	
P. ameliea ameliae (Testega + Kebar +Arfak)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	***P. "AMELIAEBIRDSHEAD"***
P. ameliea ameliae (Testega)	No	No	No	No	No	No	No	No	NA	
P. ameliea ameliae (Kebar)	No	No	No	No	No	No	No	No	NA	
P. ameliea ameliae (Arfak)	No	No	No	No	No	No	No	No	NA	
P. ameliae (Nabire + OkSibil)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	***P."WEYLANDENSIS&OKSIBIL"***
P. ameliae weylandensis (Nabire)	Yes	NA	No	No	No	No	No	NA	Yes	
P. ameliae ameliae (OkSibil)	Yes	NA	No	No	No	No	No	NA	Yes	

FIGURES

Figure 1. Phylogenetic relationships as recovered in the MrBayes and RAxML analyses of the concatenated nextRAD dataset. Color boxes at each node are bootstrap support value of the RaxML analysis and posterior probability of the MrBayes analysis. Tip labels indicate species followed by location.

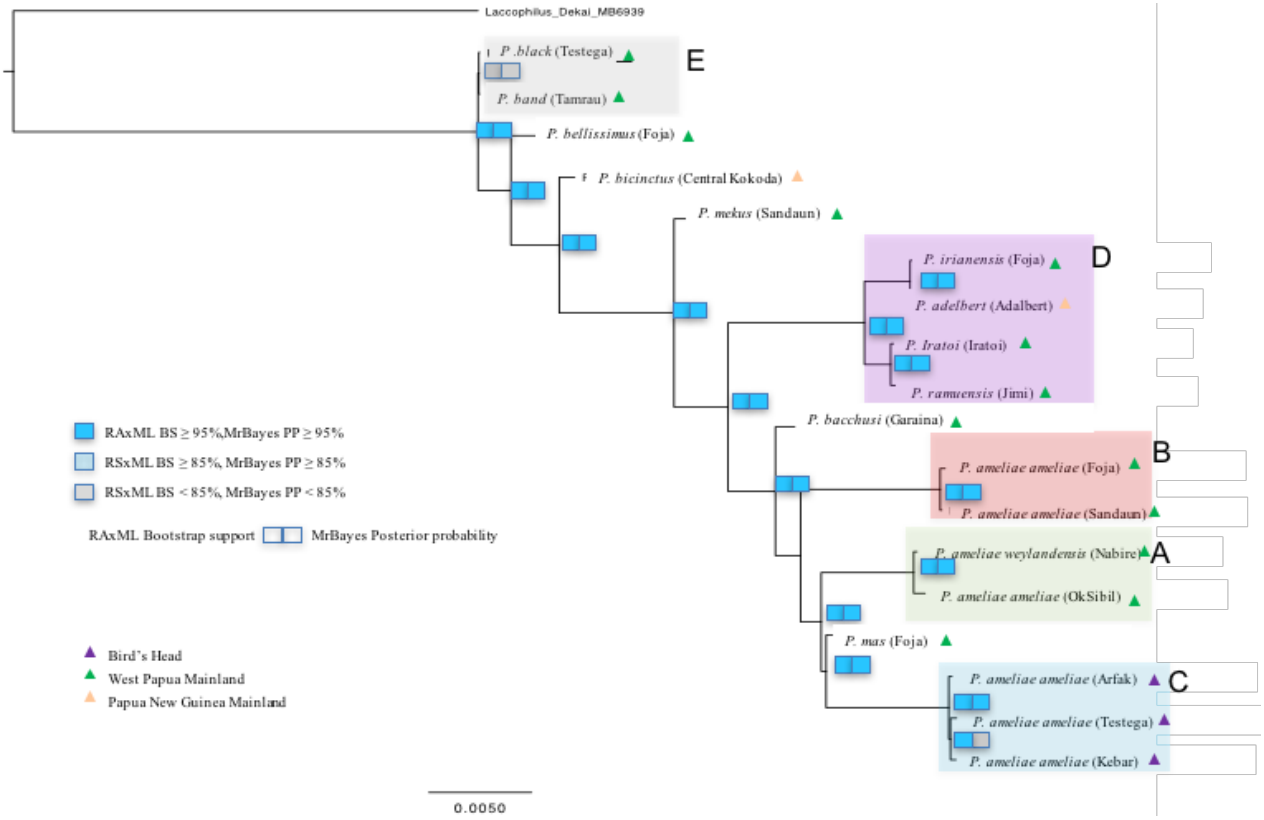


Figure 2. Species tree as recovered in ASTRAL. Color boxes at each node represents posterior probability.

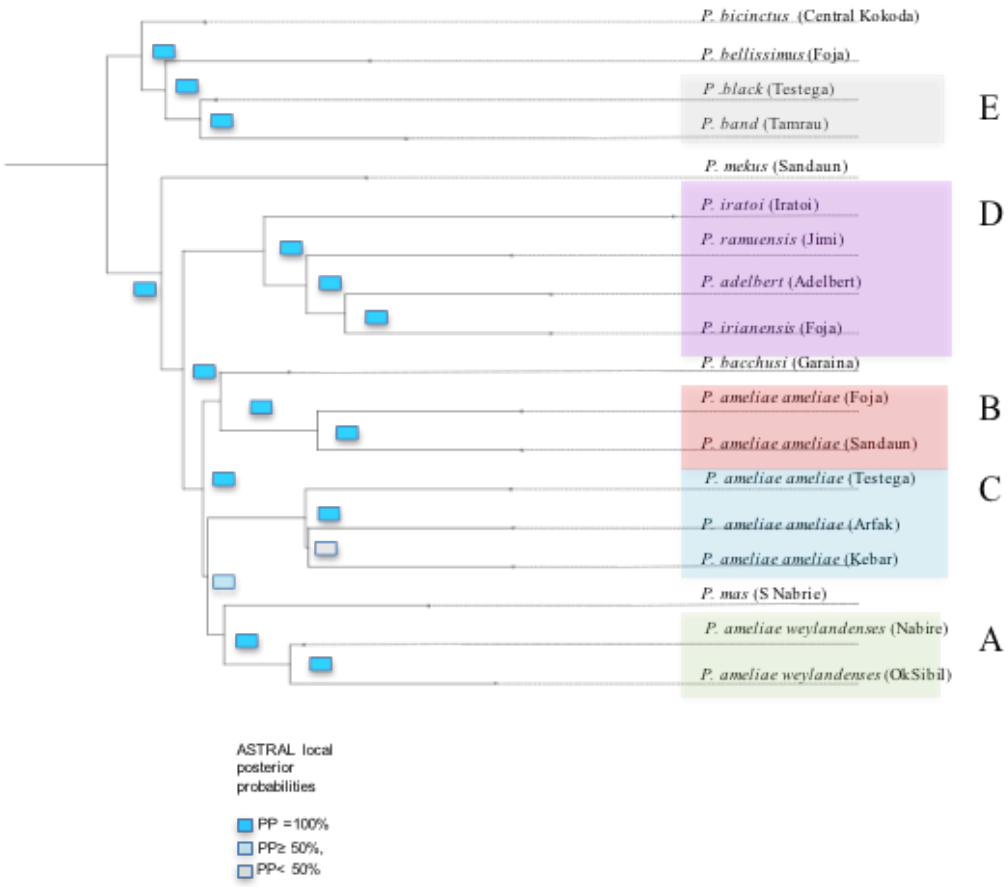


Figure 3. Species tree recovered in *SVDquartets*.

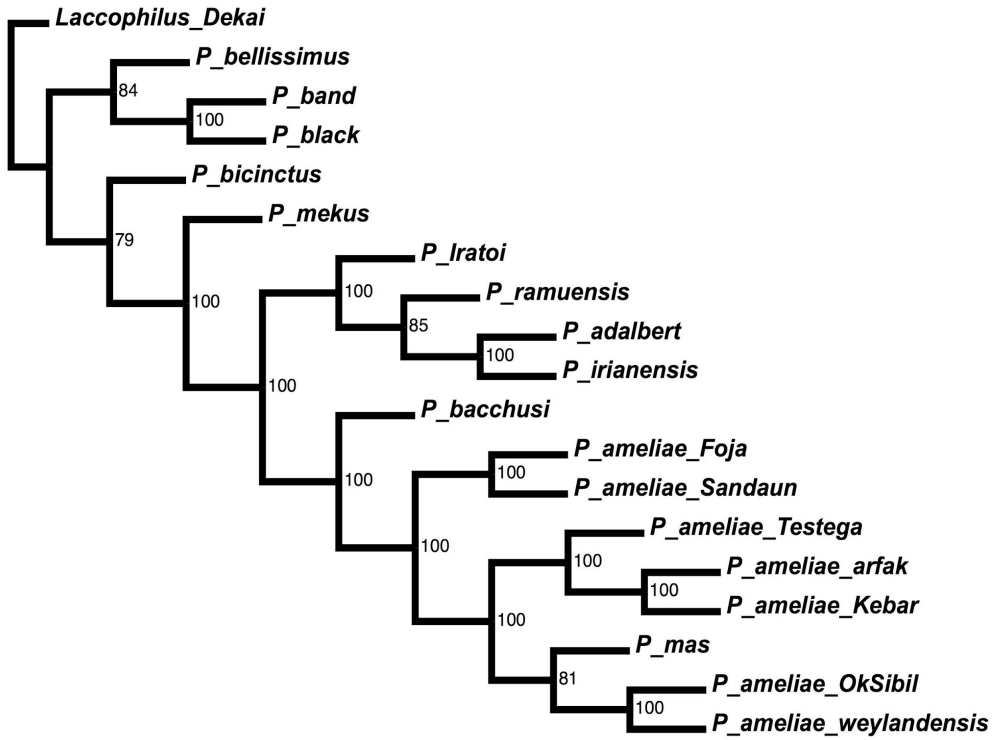


Figure 4. (Bayesian clustering analyses of nextRAD SNPs data in STRUCTURE. Barplots from STRUCTURE runs for (top) all samples (*K=1-10*) and (bottom) the genetic clusters A, -E separately (*K* = 2 for each successive run)

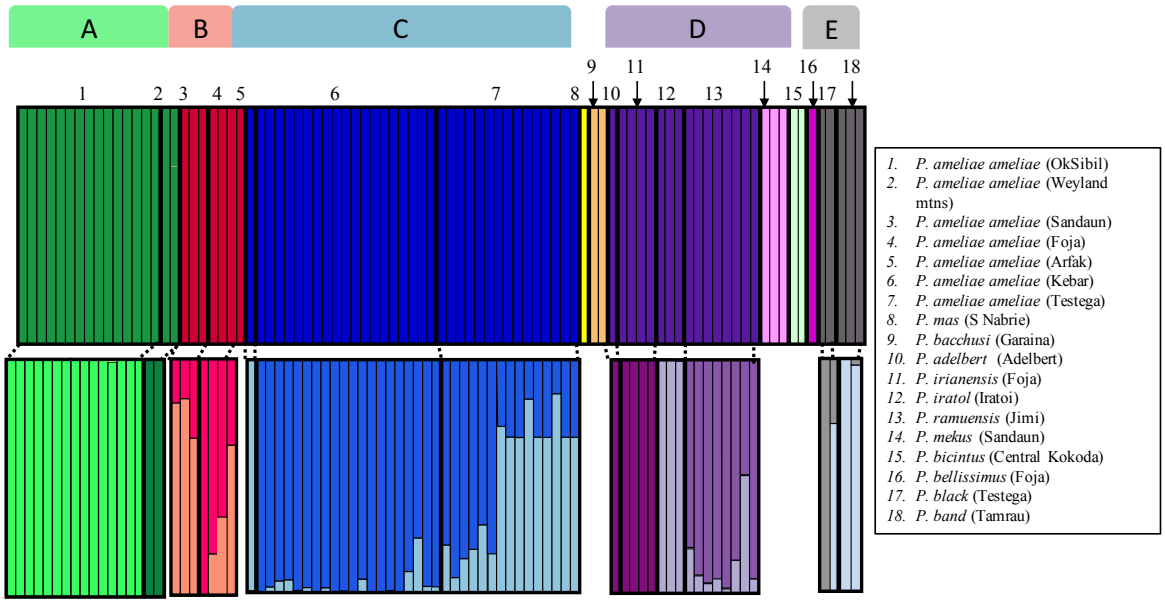
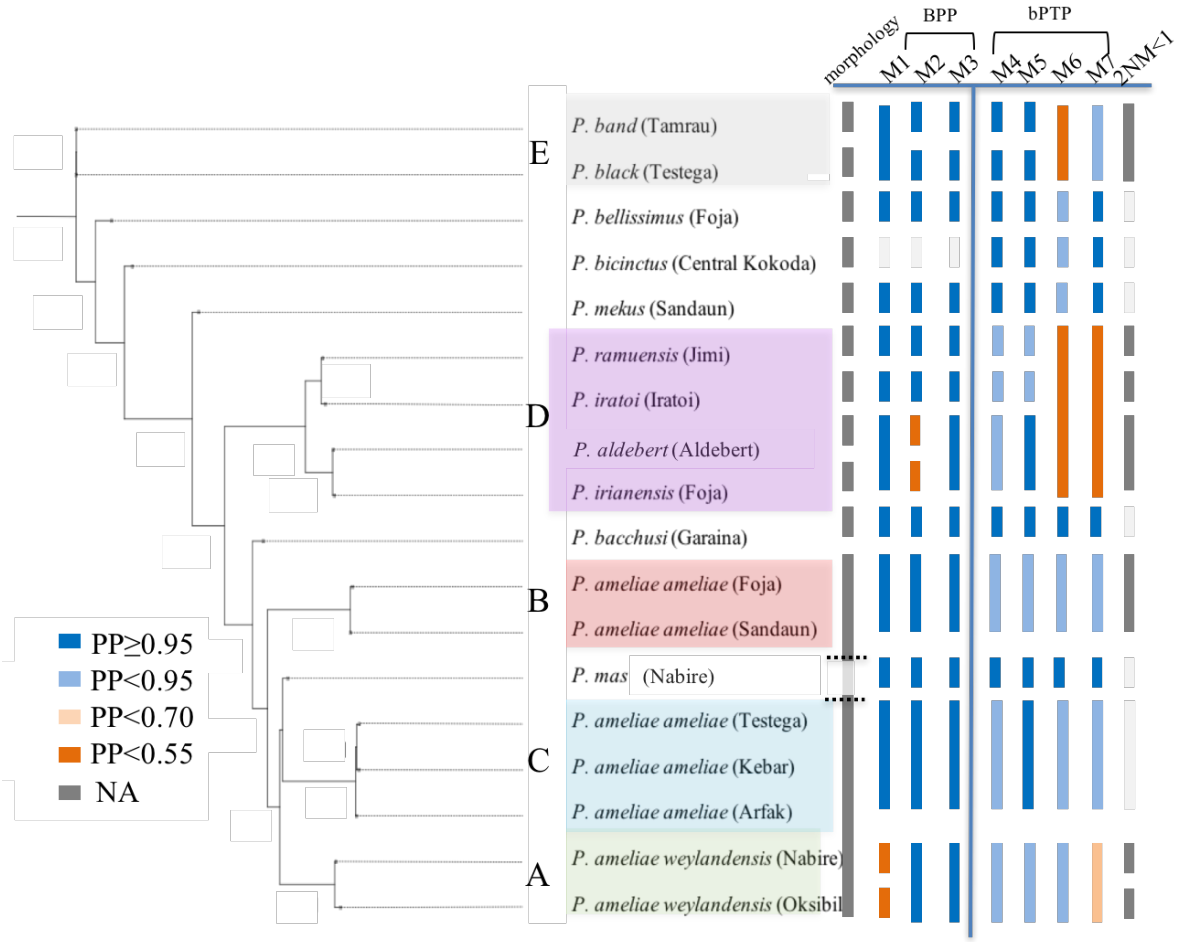


Figure 5. Species delimitation results. Each column represents results from particular method (M1-M7, see table 1). Continuous bars indicate clumping of OTU. Color of bars indicates support (posterior probability).



Supplemental Material

Table S1. Taxon sampling used in this study

Genus	Species	Code	Country	Province	Locality
Philaccolilus	*ameliae*	MB6883	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6884	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6885	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6886	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6887	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6890	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6891	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6892	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6893	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6894	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6895	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6896	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB6897	Indonesia	Papua Barat	above Kebar, open river
Philaccolilus	*ameliae*	MB3053	Indonesia	West Papua	Arfak
Philaccolilus	*ameliae*	MB6461	Indonesia	Papua	Foja
Philaccolilus	*ameliae*	MB6462	Indonesia	Papua	Foja
Philaccolilus	*ameliae*	MB6851	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6852	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6853	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6854	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6855	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB6856	Indonesia	Papua Barat	Kebar Valley
Philaccolilus	*ameliae*	MB3735	PNG	Sandaun	Mianmin
Philaccolilus	*ameliae*	MB3736	PNG	Sandaun	Mianmin
Philaccolilus	*ameliae*	MB4939	PNG	Sandaun	Mianmin area
Philaccolilus	*ameliae*	MB4940	PNG	Sandaun	Mianmin area
Philaccolilus	*ameliae*	MB4207	Indonesia	Papua	Road Nabire-Enarotali KM 95
Philaccolilus	*ameliae*	MB4208	Indonesia	Papua	Road Nabire-Enarotali KM 95
Philaccolilus	*ameliae*	MB6924	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6925	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6926	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6927	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6928	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6929	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6930	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6931	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv

Philaccolilus	*ameliae*	MB6932	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6933	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6934	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6935	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6936	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6937	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6938	Indonesia	Papua	S Ok Sibil, tributary of Digul Riv
Philaccolilus	*ameliae*	MB6857	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6858	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6898	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6899	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6900	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6901	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6902	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6903	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6904	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6905	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6906	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6907	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6908	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6909	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6910	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6911	Indonesia	Papua Barat	Testega
Philaccolilus	*ameliae*	MB6912	Indonesia	Papua Barat	Testega
Philaccolilus	*bacchusi*	MB3835	PNG	Morobe	Garaina
Philaccolilus	*bacchusi*	MB3838	PNG	Morobe	Garaina
Philaccolilus	*bellissimus*	MB6464	Indonesia	Papua	Sarmi, Waaf, N Foja Mts, riverbank
Philaccolilus	*bicinctus*	MB4116	PNG	Central	Kokoda Trek
Philaccolilus	*bicintctus*	MB2845	PNG	National Capital District	Varirata NP
Philaccolilus	*black*	MB6227	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream
Philaccolilus	*black*	MB6228	Indonesia	Papua Barat	Tamrau Mts N of Kebar, forest stream
Philaccolilus	*black band*	MB6915	Indonesia	Papua Barat	Testega
Philaccolilus	*Incognitus*	MB4087	PNG	Madang	Highway nr Madang, ford
Philaccolilus	*Iratoi*	MB6916	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6917	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6918	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6919	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6920	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6921	Indonesia	Papua	S Iratoi, forest

Philaccolilus	*Iratoi*	MB6922	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*Iratoi*	MB6923	Indonesia	Papua	S Iratoi, forest
Philaccolilus	*irianensis*	MB3046	PNG	Sandaun	Bewani Stn.,base of Bewani Mts
Philaccolilus	*irianensis*	MB5139	Indonesia	Papua	Road Nabire-Enarotali KM 108
Philaccolilus	*irianensis*	MB6459	Indonesia	Papua	Sarmi, Waaf, N Foja Mts, riverbank
Philaccolilus	*irianensis*	MB6460	Indonesia	Papua	Sarmi, Waaf, N Foja Mts, riverbank
Philaccolilus	*mas*	MB3043	Indonesia	Papua	Road Nabire - Ilaga, KM 117
Philaccolilus	*mekus*	MB4936	PNG	Sandaun	Mianmin area
Philaccolilus	*mekus*	MB4937	PNG	Sandaun	Mianmin area
Philaccolilus	*mekus*	MB6025	PNG	Sandaun	Mianmin area
Philaccolilus	*ramuensis*	MB2871	PNG	Western Highlands	Jimi Valley, Sendiap Station
Philaccolilus	*ramuensis*	MB6400	PNG	Western Highlands	Jimi Valley, Sendiap Station
Philaccolilus	*ramuensis*	MB6401	PNG	Western Highlands	Jimi Valley, Sendiap Station

Note: PNG, Papua New Guinea; KM, kilometer; Mts, mountains; S, South.

Figure S1. Phylogeny based on 5068 bps of Sanger sequencing dataset. Four approaches and parameter choices were used: IQtree, RAxML, MrBayes, MrBayes reversible -jump. Color boxes represent support values

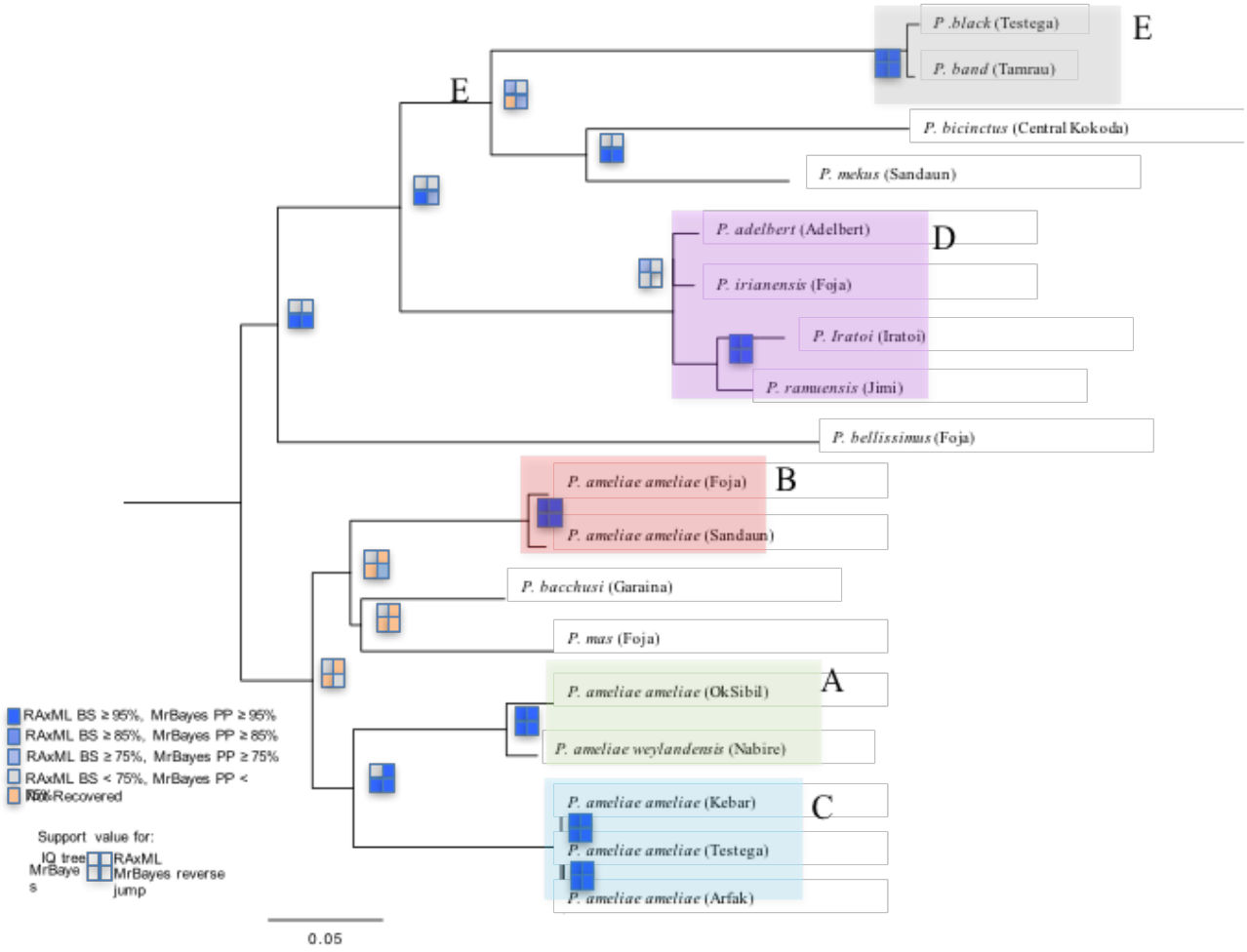


Figure S2. Dated phylogeny based on CO1 dataset recovered in BEAST